

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2003, 09:30:20 ; Search time 8148 Seconds

(without alignments)
11417.341 Million cell updates/sec

Title: US-09-494-297-1

Perfect score: 2274

Sequence: 1 atgaaaaaacaaggttcc.....ggataagaacaatgactag 2274

Scoring table: IDENTITY_NUC

Gap 10.0 , Gape 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

GenEmbl:*

- 1: gb.ba:*
- 2: gb.htg:*
- 3: gb.in:*
- 4: gb.om:*
- 5: gb.ov:*
- 6: gb.pat:*
- 7: gb.ph:*
- 8: gb.pl:*
- 9: gb.pr:*
- 10: gb.ro:*
- 11: gb.sts:*
- 12: gb.sy:*
- 13: gb.un:*
- 14: gb.vi:*
- 15: em.ba:*
- 16: em.fun:*
- 17: em.hum:*
- 18: em.in:*
- 19: em.mu:*
- 20: em.om:*
- 21: em.or:*
- 22: em.ov:*
- 23: em.pat:*
- 24: em.ph:*
- 25: em.pl:*
- 26: em.ro:*
- 27: em.sts:*
- 28: em.un:*
- 29: em.vi:*
- 30: em.htg.hum:*
- 31: em.htg.in:*
- 32: em.htg.other:*
- 33: em.htg.mu:*
- 34: em.htg.pln:*
- 35: em.htg.ro:*
- 36: em.htg.mam:*
- 37: em.htg.vrt:*
- 38: em.sy:*
- 39: em.htgo.hum:*
- 40: em.htgo.mu:*
- 41: em.htgo.other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2274	100.0	11289	1	AE006482	AE006482 Streptococcus
2	2270.8	99.9	2400	1	AB083107	AB083107 Streptococcus
3	756.4	33.3	16207	1	AF447492	AF447492 Streptococcus
4	720.8	31.7	10826	1	SPU43397	SPU43397 Streptococcus
5	664.4	29.2	52900	1	AE014138	AE014138 Streptococcus
6	664.4	29.2	311600	1	AP005141	AP005141 Streptococcus
7	438.4	19.3	11420	1	AE009963	AE009963 Streptococcus
8	215.8	9.5	2845	1	STREPINF	STREPINF Streptococcus
9	180.2	7.9	2263	1	SPSFB	SPSFB Streptococcus
10	167.6	7.4	2066	1	GGU31115	GGU31115 Streptococcus
11	150.2	6.6	2801	1	AF009908	AF009908 Streptococcus
12	97	4.3	321	1	AF009914	AF009914 Streptococcus
13	92.8	4.1	313	1	AF009913	AF009913 Streptococcus
14	75.8	3.3	293650	1	AP004172	AP004172 Streptococcus
15	73	3.2	145265	2	AL935272	AL935272 Danio rerio
16	72	3.2	348	1	AF009920	AF009920 Streptococcus
17	71	3.1	110000	2	PFMAL13_07	PFMAL13_07 Streptococcus
18	69.8	3.1	155204	2	AC007926	AC007926 Trypanosoma
19	68.6	3.0	203138	2	BX530070	BX530070 Danio rerio
20	68.4	3.0	309	1	AF009915	AF009915 Streptococcus
21	68.2	3.0	278	1	AF009917	AF009917 Streptococcus
22	68	3.0	250029	3	AE014820	AE014820 Plasmodium
23	67.6	3.0	281	1	AF009909	AF009909 Streptococcus
24	67.6	3.0	282	1	AF009912	AF009912 Streptococcus
25	66.8	2.9	310	1	AF009911	AF009911 Streptococcus
26	66.8	2.9	110000	3	AC116984_3	AC116984_3 Streptococcus
27	66	2.9	94434	5	AL929250	AL929250 Zebrafish
28	65.6	2.9	148871	2	AL929469	AL929469 Danio rerio
29	65.6	2.9	213544	2	BX510939	BX510939 Danio rerio
30	65.6	2.9	225581	2	BX537105	BX537105 Danio rerio
31	65	2.9	110000	2	BX005010_3	BX005010_3 Streptococcus
32	65	2.9	198550	2	BX004991	BX004991 Danio rerio
33	64.2	2.8	326	1	AF009910	AF009910 Streptococcus
34	64.2	2.8	2642	3	AF103869	AF103869 Plasmodium
35	63.4	2.8	273275	3	AE014828	AE014828 Plasmodium
36	63.4	2.8	330050	3	PR929355	PR929355 Plasmodium
37	62.8	2.8	250029	3	AE014816	AE014816 Plasmodium
38	61.6	2.7	309	1	AF009918	AF009918 Streptococcus
39	61.6	2.7	185730	2	BX005453	BX005453 Danio rerio
40	61.4	2.7	255	1	AF009916	AF009916 Streptococcus
41	60.8	2.7	110000	2	PFMAL13_24	PFMAL13_24 Streptococcus
42	60.6	2.7	110000	2	PFMAL13_12	PFMAL13_12 Streptococcus
43	60.6	2.7	110000	2	PFMAL13_13	PFMAL13_13 Streptococcus
44	60.6	2.7	333321	3	AC116986	AC116986 Dicycloste
45	60.4	2.7	719	3	AY277504	AY277504 Plasmodium

ALIGNMENTS

RESULT 1
LOCUS AE006482
DEFINITION Streptococcus pyogenes M1 GAS strain SF370, section II of 167 of the complete genome.
ACCESSION AE006482 AE004092
VERSION AE006482.1 GI:13621422
KEYWORDS
SOURCE
ORGANISM Streptococcus pyogenes M1 GAS
Streptococcus pyogenes M1 GAS
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 11289)
Ferretti J.J., McShan W.M., Adji D., Savic D., Savic G., Lyon K., Primeaux C., Szatek S.S., Surorov A.N., Kenton S., Lai H., Lin S.,

Qian, Y., Jia, H. G., Najjar, F. Z., Ren, Q., Zhu, H., Song, L., White, J.,
Yuan, X., Clifton, S. W., Roe, B. A. and McLaughlin, R. E.
Complete genome sequence of an M1 strain of *Streptococcus pyogenes*
Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)

JOURNAL
PUBMED
21192684
11296296

AUTHORS
REFERENCE
2 (bases 1 to 11289)
Ferretti, J. J., McShan, W. M., Ajdic, D., Savic, D., Savic, G., Lyon, K.,
Primeaux, C., Sezate, S. S., Surorov, A. N., Kenton, S., Lai, H., Lin, S.,
Qian, Y., Jia, H. G., Najjar, F. Z., Ren, Q., Zhu, H., Song, L., White, J.,
Yuan, X., Clifton, S. W., Roe, B. A. and McLaughlin, R. E.
Direct Submission
Submitted (10-APR-2001) Department of Microbiology and Immunology,
University of Oklahoma Health Sciences Center, 940 St. Young Blvd.,
Oklahoma City, OK 73104, USA

FEATURES
SOURCE
1. 11289
Location/Qualifiers
/organism="Streptococcus pyogenes M1 GAS"
/mol_type="genomic DNA"
/strain="SF370"
/serotype="M1"
/db_xref="taxon:160490"
complement(1496..2137)
/gene="SpY0121"
complement(1496..2137)
/gene="SpY0121"
/EC_number="2.7.1.113"
/note="Best Blastp hit = dbj|BAB16044.1| (AB030884)
similar to Lactobacillus acidophilus da/dg-kinase
[Streptococcus zooepidemicus]
/codon_start=1
/transl_table=11
/product="putative deoxyguanosine kinase/deoxyadenosine
kinase(I) subunit"
/protein_id="AAK33233.1"
/db_xref="GI:13621423"
/translation="MLIVLAGTIGAGKSSIAALGHELTGDFVEAVDNNPVLDTLYO
DPKRYAFLLQIVFLNKRKFSIKAYQADNNILDRSIFEDFLKLNKNNVTKELD
IYQELANMLEELGMPKRPDLIYIDVSEFMLEIERGRSFEQVGNPSLEQY
HQVGEPTVYEDYEVSPKMKIDGNSLDFVQNPDLATVLMKIDTKLEHLL"
complement(2157..3134)
/gene="SpY0122"
complement(2157..3134)
/gene="SpY0122"
/note="Best Blastp hit = gb|AAC97155.1| (U49397) unknown
[Streptococcus pyogenes]"
/codon_start=1
/transl_table=11
/product="possible transcriptional regulator"
/protein_id="AAK33234.1"
/db_xref="GI:13621424"
/translation="MIRKNSFRIGDVELPRTVLAPMAGVTSAPFTIATKEGAGLY
VMDISEKGLYNNKRTLHMLHIDENHPMSIOFGDAGLKRADFIOINRKADIV
DINWGPVNVKVAENAGAKVLRDPRDKYHIVKVTSTGLDPLTKMRTGADSLAVE
NALAASAGYASALAMHGRPREOMYTGCDHEETLARSKATKTIPTGNGRVSVOAK
FMLEIEGYDAVMIGRAAMNPTLFTQINHEFTFGQELPDLPPAKKLDIAODHLKRLIN
LKGETIAVREFRLAPHYLRGTAGAAVRAVSAETLAEVAIFETVR"
complement(3121..3993)
/gene="SpY0123"
complement(3121..3993)
/gene="SpY0123"
/note="Best Blastp hit = dbj|BAB03806.1| (AP001507)
unknown conserved protein [Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="putative heat shock protein HSP33"
/protein_id="AAK33235.1"
/db_xref="GI:13621425"
/translation="MDKIISIAQSGAFRAVYVDSTETVALAQEKHNTLSSSTVALGR
TLIANQILANQKSGSKIVKIVIGSDSFGHIIISVADTKGVKGYIONTGVDIKRTATG
EVLVGPFGNGHFTIIDGNGPRTSTPLTIGEDFAVYLTSEQCPISGLVAV
LIDENKVKVAGGFVQVILPGASSEERLAEKLOMPAISHLASKNHVDLLEIY
GDPRYKRLSEPTSPQDCSRENFEPALMTLPKADQAMIDEDKGAETIVCGFGTKTKO

FNESDLFAISDKA"
complement(4140..5633)
/gene="rofa"
/note="synonym: SpY0124"
complement(4140..5633)
/gene="rofa"
/note="Best Blastp hit = gb|AAC43219.1| (U01312) RoFA
[Streptococcus pyogenes]"
/codon_start=1
/transl_table=11
/product="regulatory protein"
/protein_id="AAK33236.1"
/db_xref="GI:13621426"
/translation="MIEKYLESIESKQILVLPFKTSYLPITEVAEKTLFELQLNH
YCEELNAPFGSLSMITOKRMISQGFHPKREYLYQLVASNVLOLAFILKNGSH
RPLTDFARSHPLSNSAYRMRREALIPLRPELKISKNTKIVGEERYIRYLIALLSFE
GIKYVDLTQODKNTIHSFLSHSTHLKTSWLSSESFSFYDILALSKRRQFSYTTIQ
TRIFQQLKRLVYDSLRKSSHDIIETVQQLNFSAGDLVLYLYITANNFSASLOWTP
EHIROYQLEENDTERLNLNPIITLPLNLEQKASLVKALMFESKSFLENLOHFIPE
TNLFVSPYKGNOKLYSLKLIYENMAKLPGRDLNKKHFLCHVQEOSLRNIOFP
LVVFEVANSFINAHILTDSPRYSKSIDFHSYLLQDNVYOIPDLKPDLYVTHSOL
IPVYHETLTGIAVAELSPDESILSTQELMYQVKEEKFQADLRKQL"
8324..8719
/gene="SpY0127"
8324..8719
/note="Best Blastp hit = gb|AAC97149.1| (U49397) LepA
[Streptococcus pyogenes]"
/codon_start=1
/transl_table=11
/product="putative signal peptidase I"
/protein_id="AAK33237.1"
/db_xref="GI:13621427"
/translation="MIKRNMDAPSVKAGDAILEYRLSOTYKVEAVVYEDSSTIK
VGRITQAGDEVDLTGEGELKINGHONBELTITKSEANYPPIRIADNSYLLINDIYS
QSESNITQDAIAKDAIKGTINTLIRLNH"
8752..9774
/gene="SpY0128"
8752..9774
/note="Best Blastp hit = gb|AAC97152.1| (U49397) unknown
[Streptococcus pyogenes]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK33238.1"
/db_xref="GI:13621428"
/translation="MKIRHLTLTGALTSRAATVGHGVNGAKLTVTKNLDVNSN
ALIPNTDFTKIEPDTTVNEDGKRGVALNTPTKYTTINSRSGSNRTTAEDFSE
VTEKPGVYIKVTEKIDKVPVSVSDTTSYVQVHVMNEQKRFVAYIVGKESG
KVIDIQRKNSLSTTLVYKRVSGTGDGRSKDFNGLKANQYKASEKVMIEKFG
GQAPVQTEASIDQTLVYTKLDGDSIVVTNLPVQVYVEDDYKSKRYTNNVSVSD
GAVNKIAGNSETEQSTDKDMTITFNKDFEVPQVAMTVABYIALGIVAVGALYF
VKKRNA"
9784..10497
/gene="SpY0129"
9784..10497
/note="Best Blastp hit = gb|AAC97152.1| (U49397) unknown
[Streptococcus pyogenes]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK33239.1"
/db_xref="GI:13621429"
/translation="MTVRIKILDKLINVIVICFPFLCILIALGTYDALTYYOGANA
TNQOYKRRKQVQFDLLAINSQVMAITVKGHTIDYPIYQGENNLEIKNSVGEISL
SGSVFLDYRKRVTEDEKYSLLTAHMGAGNREDELNFRKFSFNNHKEISLETCKO
KLKINIFACIOTDAFDSLLENPIIDVIVISSKNEFLNIHKOSVQYREILTTNESRVAL
STCEDMTGRIIRIVGQIE"
10519..11166
/gene="SpY0130"
10519..11166

QY 1921 GTTAAACAGATTAACAAACCTCGAATTAAAGTGTAAAGCAACCATTAATTAA 1980
|||||
DB 7816 GTTAAACAGATTAACAAACCTCGAATTAAAGTGTAAAGCAACCATTAATTAA 7875
1981 CATGGGGAAGTTTAACTTCAAGGTTTACCAGAGTTTATCTTACCTGTCAAGAA 2040
|||||
DB 7876 CATGGGGAAGTTTAACTTCAAGGTTTACCAGAGTTTATCTTACCTGTCAAGAA 7935
QY 2041 ACAATTTCTGAAGCTTAAAGTTAAAGTAAATAGCCAGAGTAAATGCTACAGT 2100
|||||
DB 7936 ACAATTTCTGAAGCTTAAAGTTAAAGTAAATAGCCAGAGTAAATGCTACAGT 7995
QY 2101 TCAAAAACAGAAATACAGATGATGACACTGTGTTGAAATTAATTAAGAGCTGT 2160
|||||
DB 7996 TCAAAAACAGAAATACAGATGATGACACTGTGTTGAAATTAATTAAGAGCTGT 8055
QY 2161 GTTCTTACAGAGATGATCAAAAGATCAATGCTATCTAGCTTGTATGATGCTGTG 2220
|||||
DB 8056 GTTCTTACAGAGATGATCAAAAGATCAATGCTATCTAGCTTGTATGATGCTGTG 8115
QY 2221 ATCAGTTGGGATCTGGGGAATTCACACGATTAAGATTAAGAAACATGACTAG 2274
|||||
DB 8116 ATCAGTTGGGATCTGGGGAATTCACACGATTAAGATTAAGAAACATGACTAG 8169

RESULT 2
AB083107 2400 bp DNA linear BCT 21-JAN-2003
LOCUS Streptococcus pyogenes cpai gene for collagen binding protein,
DEFINITION complete cds.
ACCESSION AB083107 GI:23503477
VERSION AB083107.1
KEYWORDS
SOURCE Streptococcus pyogenes
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus pyogenes

REFERENCE
AUTHORS Miyoshi-Akiyama,T., Wakisaka,N., Zhao,J. and Uchiyama,T.
TITLE Characterization of collagen binding protein (Cpa) of M1-type
group A streptococcus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2400)
AUTHORS Miyoshi-Akiyama,T., Wakisaka,N., Zhao,J. and Uchiyama,T.
TITLE Direct Submission
JOURNAL Submitted (03-APR-2002) Tohru Miyoshi-Akiyama, Tokyo Women's
Medical University, Department of Microbiology and Immunology; 8-1,
Kanda-cho, Shinjuku-ku, Tokyo 162-8666, Japan
(E-mail:KFA01112@nifty.ne.jp, Tel:81-3-3353-8111(ex.22713),
Fax:81-3-5269-7411)
FEATURES
source location/Qualifiers

1..2400
/organism="Streptococcus pyogenes"
/mol_type="genomic DNA"
/strain="291"
/serotype="M1"
/db_xref="taxon:1314"
/country="Japan"
94..2367
/gene="cpai"
94..2367
/gene="cpai"
/transl_table=1
/codon_start=1
/product="collagen binding protein"
/protein_id="BAC20340.1"
/db_xref="GI:23503478"
/translation="MKTRPNNKLTINTQVLSKSKRFTVLGVLMFALVTSM
VGAKTVGLVSSSTPNAIIPDSSEYRGVYKOPRVAHDLRVNLSGR
SYOYCNKIKKAPLIGSDSSYKKKKHDSKPEYKAISSPTGELNKLAAVY
NGHPQNGKIRIEGLPLAIFVTOEAVWYSNDAPISNDSFRSEESINVTSSLS
LMROALKQLIDPNIATKMPKOVDPDQLSIFESBDDKDKYGNLLSGLVPTKRP

TPDPPMPNPOTTSVLRKRYAGDYSKLLEGATLOLTDGNVNSFOARFSSNDIGE
RIELSDCTVTLLELNSPAGYSIAPRIEFKQYAGVYVILIDGKOLENKNKELEYSVE
AYNDPEERSVLTITONVAKFYAKKNSSSOYVYFVNDIKSPDESEGGCTMPDPTT
GEVATYTHIAGDLEKYVAKPRDHPDFFLHKIRKVIKGRKREGQALEYSGLETOLR
AAIOLAIFYFDSDSELDDKIKIDHGGDNMDSYLAIAKILVEYAQDSNPOLDLF
FIPNNKRYOSLIGTQWHPEDLVLDIIRBEDKKEVLPVHAKILKRTVYGLAGDRKDFH
FEILKNKNOELISQVTKETNLEFDGKATILKKGESLTLQGLEEGSYLVKED
SEGKRVKNVSOEVNATVSKTGISDSTLAFENKKEVYVPTGVQDKINGLIALIVAG
ISLGIWGHITRIKRD"
BASE COUNT 873 a 364 c 463 g 700 t
ORIGIN
Query Match 99.9%; Score 2270.8; DB 1; Length 2400;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2272; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGAAAAAACAGGTTTCCAAATTAAGTTAACTCTTAATTAAGGTTTAACT 60
|||||
DB 94 ATGAAAAAACAGGTTTCCAAATTAAGTTAACTCTTAATTAAGGTTTAACT 153
61 AAAAAGTCAAAAGATTACTGTCAGTTAGTGGAGCTTTTAAATGATGCTGCTTG 120
|||||
DB 154 AAAAAGTCAAAAGATTACTGTCAGTTAGTGGAGCTTTTAAATGATGCTGCTTG 213
121 GTAACCTCCAGTGTGGTGAAGACGCTTTTGTAGTAGAATCCTGACGCCAAC 180
|||||
DB 214 GTAACCTCCAGTGTGGTGAAGACGCTTTTGTAGTAGAATCCTGACGCCAAC 273
QY 181 GCATTAATCCAGATTCAAGTTGCGAATACAGATGATATGATATGATCTTAAGTA 240
|||||
DB 274 GCAATTAATCCAGATTCAAGTTGCGAATACAGATGATATGATATGATCTTAAGTA 333
241 GGGCATCCATTTTAAACAGATTAGAGTACACAGATTTAAAGGTTAACTTGAAGA 300
|||||
DB 334 GGGCATCCATTTTAAACAGATTAGAGTACACAGATTTAAAGGTTAACTTGAAGA 393
394 AGTAGAAGTTATCAAGTTTATGCTTAATTAAGAAAGCATTTCCCTCGAGTACAG 453
QY 301 AGTAGAAGTTATCAAGTTTATGCTTAATTAAGAAAGCATTTCTCGAGTACAG 360
|||||
DB 394 AGTAGAAGTTATCAAGTTTATGCTTAATTAAGAAAGCATTTCCCTCGAGTACAG 453
QY 361 AGTAGAAGTTATCAAGTTTATGCTTAATTAAGAAAGCATTTCTCGAGTACAG 420
|||||
DB 454 AGTAGAAGTTATCAAGTTTATGCTTAATTAAGAAAGCATTTCTCGAGTACAG 513
421 GCGATGAGCCCTTAAGTTTACGGAGATGAGCTTAATCAAGAGTACAGCTGTATAT 480
|||||
DB 514 GCGATGAGCCCTTAAGTTTACGGAGATGAGCTTAATCAAGAGTACAGCTGTATAT 573
QY 481 AATGACATCCACAAATGCCAATGATTAATGGAAGGCTTGAACCCCTGATGATC 540
|||||
DB 574 AATGACATCCACAAATGCCAATGATTAATGGAAGGCTTGAACCCCTGATGATC 633
541 AGAGTTACACAGAGGGGGTATGCTATCTGATATATAGTCTTATTAATCCAGAT 600
|||||
DB 634 AGAGTTACACAGAGGGGGTATGCTATCTGATATATAGTCTTATTAATCCAGAT 693
QY 601 GAAAGTTTAAAGGAGTCAAGAAATGCTTGTAGTACTTCTCAATTAATCTTGTATG 660
|||||
DB 694 GAAAGTTTAAAGGAGTCAAGAAATGCTTGTAGTACTTCTCAATTAATCTTGTATG 753
QY 661 CGTCAAGCTTGAAGCAACGATGATCCGAAATTTGCAACTTAATTAATTAATTA 720
|||||
DB 754 CGTCAAGCTTGAAGCAACGATGATCCGAAATTTGCAACTTAATTAATTAATTA 813
721 CCGGATGATTTGAGCTAAGATTTTGAAGCTGAGCAAGGAGATTAATTAATTA 780
|||||
DB 814 CCGGATGATTTGAGCTAAGATTTTGAAGCTGAGCAAGGAGATTAATTAATTA 873
QY 814 GGAATCAAAATCTTTGAGTGTGCTTACTCTTAATCAACCAACCTCCAGAGAC 840
|||||
DB 874 GGAATCAAAATCTTTGAGTGTGCTTACTCTTAATCAACCAACCTCCAGAGAC 933
QY 841 CCACCAATGCTCCAAATCAACCTCAACGACTTCACTTATTAAGAAAGTATGCTATA 900

Db	934	CCACCAATGCTCCAAATCAACCTCAACGACTCAGTACTTATTAGAAAGTATGCTATTA	993
Qy	901	GGTGATTAATCTAAATTCCTTGAAGTGCAACATTACAGTTGACAGGGGATPACGTGAAT	960
Db	994	GGTGATTAATCTAAATTCCTTGAAGTGCAACATTACAGTTGACAGGGGATPACGTGAAT	1053
Qy	961	AGTTTTCAGAGGAGAGTGTTTAGACATATATATGAGAGAAAGATTTGAATCATGAT	1020
Db	1054	AGTTTTCAGAGGAGAGTGTTTAGACATATATATGAGAGAAAGATTTGAATCATGAT	1113
Qy	1021	GGAACTTACTTTTAACTGAATTCCTCAGCTGGTATPAGATGSCAGAGCAATC	1080
Db	1114	GGAACTTACTTTTAACTGAATTCCTCAGCTGGTATPAGATGSCAGAGCAATC	1173
Qy	1081	ACTTTTAAAGTTGAAGCTGGCAAAAGTATATATATTTATGAGAAACAGATTTGAAT	1140
Db	1174	ACTTTTAAAGTTGAAGCTGGCAAAAGTATATATATTTATGAGAAACAGATTTGAAT	1233
Qy	1141	CCCAATTAAGAGATATAGTACCTTACTCAGTACAGCATATTAATGATTTGAGAAATTT	1200
Db	1234	CCCAATTAAGAGATATAGTACCTTACTCAGTACAGCATATTAATGATTTGAGAAATTT	1293
Qy	1201	AGCGTTTAACTACACAAACTATGCAAAATTTTATATGCAAAATTAATAATGCAAGT	1260
Db	1294	AGCGTTTAACTACACAAACTATGCAAAATTTTATATGCAAAATTAATAATGCAAGT	1353
Qy	1261	TCACAGGTTGTCTATTCCTTAAATGACAGATCTAAATTCACACAGACTGTGAAGATGT	1320
Db	1354	TCACAGGTTGTCTATTCCTTAAATGACAGATCTAAATTCACACAGACTGTGAAGATGT	1413
Qy	1321	GGGAAACAAATGACTCCAGACTTTTACACAGAGAGAAATATACATCATATTCAGAGT	1380
Db	1414	GGGAAACAAATGACTCCAGACTTTTACACAGAGAGAAATATACATCATATTCAGAGT	1473
Qy	1381	CGTGACCTCTTAAATATATAGTGTGAACCAAGAGATACCGATCCGACACTTTCTTAAA	1440
Db	1474	CGTGACCTCTTAAATATATAGTGTGAACCAAGAGATACCGATCCGACACTTTCTTAAA	1533
Qy	1441	CATATCAAAAAGTAATTTAGAGAGGTTTACAGGGAAAAAGACAACTATTGAGTATAGT	1500
Db	1534	CATATCAAAAAGTAATTTAGAGAGGTTTACAGGGAAAAAGACAACTATTGAGTATAGT	1593
Qy	1501	GGTCTACTGACACAAATTCGCTGGCGCTACTCAGTACGAATATATATTTTCTCAT	1560
Db	1594	GGTCTACTGACACAAATTCGCTGGCGCTACTCAGTACGAATATATATTTTCTCAT	1653
Qy	1561	AGTCTGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA	1620
Db	1654	AGTCTGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA	1713
Qy	1621	AGTCTGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA	1680
Db	1714	AGTCTGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA	1773
Qy	1681	CTAAGTGCCTGATTTCTTATTCGGAATTAACAATTAATTAATTAATTAATTAATTAAG	1740
Db	1774	CTAAGTGCCTGATTTCTTATTCGGAATTAACAATTAATTAATTAATTAATTAATTAAG	1833
Qy	1741	CAGTGCATCCAGAAAGATTTAGTATATTTGCTATGAGAAATTAATAAGATTAATA	1800
Db	1834	CAGTGCATCCAGAAAGATTTAGTATATTTGCTATGAGAAATTAATAAGATTAATA	1893
Qy	1801	CCGTGAATCTCAATTTAATTAACATTTGAGAAAAACGCTGACTGTTAGCTGTGACAACT	1860
Db	1894	CCGTGAATCTCAATTTAATTTAATTAACATTTGAGAAAAACGCTGACTGTTAGCTGTGACAACT	1953
Qy	1861	AAAGATTTCAATTTTGAATTAATTAATAATAATAATAAGAAATTCCTTTCTCAACT	1920
Db	1954	AAAGATTTCAATTTTGAATTAATTAATAATAATAATAAGAAATTCCTTTCTCAACT	2013
Qy	1921	GTTAAACAGATTAACAAACCTGCAATTTAAAGTGTAAAGCAACATTAATTTAA	1980
Db	2014	GTTAAACAGATTAACAAACCTGCAATTTAAAGTGTAAAGCAACATTAATTTAA	2073
Qy	1981	CATGGGAAAGTTTAAACACTTCAAGGTTTACCAGAGGTTATTTTACCTTGTCAAGAA	2040
Db	2074	CATGGGAAAGTTTAAACACTTCAAGGTTTACCAGAGGTTATTTTACCTTGTCAAGAA	2133
Qy	2041	ACAGATTTGAAAGGTTATTAAGTTTAACTTAATTAATTAATTAATTAATTAATTAATTA	2100
Db	2134	ACAGATTTGAAAGGTTATTAAGTTTAACTTAATTAATTAATTAATTAATTAATTAATTA	2193
Qy	2101	TCAAAACAGAAATTAACAGATGATGACACTGCTTTTGAATAATAATAAGAGCTGTT	2160
Db	2194	TCAAAACAGAAATTAACAGATGATGACACTGCTTTTGAATAATAATAAGAGCTGTT	2253
Qy	2161	GTTCTTACAGAGATTTGATCAAAAGATCAATGCTATCTAGCTTTGATAGTATGCTGCT	2220
Db	2254	GTTCTTACAGAGATTTGATCAAAAGATCAATGCTATCTAGCTTTGATAGTATGCTGCT	2313
Qy	2221	ATCAGTTGGGATCTGGGGAATTCACACGATTAAGATTAAGAAAAATGACTAG	2274
Db	2314	ATCAGTTGGGATCTGGGGAATTCACACGATTAAGATTAAGAAAAATGACTAG	2367
Db	AF447492	16207 bp DNA linear BCT 07-MAR-2002	
LOCUS	AF447492		
DEFINITION	AF447492	Streptococcus pyogenes RoA (rofa), protein F (prfF.12), Cpa (cpa.12), Lepa (lepa), EftLSL.A (eftLSL.A12), EftLSL.B (eftLSL.B), MsmRL (msmRL), and protein F2 (prfF2.12) genes, complete cds; and unknown gene.	
ACCESSION	AF447492		
VERSION	AF447492.1	GI:19224132	
KEYWORDS		Streptococcus pyogenes	
SOURCE		Streptococcus pyogenes	
ORGANISM		Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.	
REFERENCE		1 (bases 1 to 16207)	
AUTHORS		Bessen,D.E. and Kalia,A.	
TITLE		Genomic localization of a T serotype locus to a recombinatorial zone encoding extracellular matrix-binding proteins in Streptococcus pyogenes	
JOURNAL		Infect. Immun. 70 (3), 1159-1167 (2002)	
MEJLINE		21843101	
PUBMED		11854196	
REFERENCE		2 (bases 1 to 16207)	
AUTHORS		Bessen,D.E. and Kalia,A.	
TITLE		Direct Submission	
JOURNAL		Submitted (12-NOV-2001) Epidemiology & Public Health, Yale University, 60 College Street, New Haven, CT 06520, USA	
FEATURES		Location/Qualifiers	
source		1..16207	
		/organism="Streptococcus pyogenes"	
		/mol_type="genomic DNA"	
		/strain="A735"	
		/serotype="M12"	
		/db_xref="taxon:1314"	
		/note="Fct region"	
		complement(147..1466)	
		/gene="rofa"	
		complement(147..1466)	
		/gene="rofa"	
		/note="transcriptional regulator"	
		/codon_start=1	
		/transl_table=1	
		/product="RoFa"	
		/protein_id="AAL86404.1"	
		/db_xref="GI:19224133"	
		/translation="WTDKRMISQFHPSEYLYLYLYASSNVLQILAFLIKGSHS RPLDFAHSFLINSASVAMREALIPLRNELKLSKKRYIGEERYRIYLALYSKR GIKYVDLQODKRNIIHSFLSHSLKSPWLSFSFYDILALSMWRHOFVYTIQ TIRFOQLKLEFYDLSLKSSDIETCYQLNFSAGDDLYLYLITANFSASLOWTP EHIRQCCLPFENDTFRLLNPITILPRLNKEQKASLVKALMFPKSLFLNQHFIPE	

gene
CDS
1872..3968
/gene="prf1.12"
1872..3968
/note="fibronection-binding protein; protein F1; SfbI"
/gene="prf1.12"
/codon_start=1
/transl_table=11
/product="protein F"
/protein_id="AAL86405.1"
/db_xref="GI:19224134"
/translation="MVSSEFARCEKMNKKFLNKEAGFLVTHKKRRFAVTLGVFF
LLACAGAGIGGVAAYADEKTVNFKSPDPDYPWGDSRGIFARYHNKVKLKS
KEYOAVFNLTFRPRPTSTNNFYKKIDGSGSASRYANPVLDENDKLEKNTL
NLYNGKSNANGFMNGIEDINALVTOMAIWYISDSAPLVNDKMKMERERKNEISE
SOVTLKREALKLIDPNLEATAANKIPSGRLNLFKSENEYONLSAEVYDDPPK
GDTSEHNKTPPELDGPIIPDPKRPDSSSEPALPPLMPELDGEEVPEVPSSELPALP
PLMPELDGEEVPEVPSSELPALPPLMPELDGEEVPEVPSSELPALPPLMPELDGEE
VPEVPSSELPALPPLMPELDGEEVPEVPSVDPDIEVPRYFNKKDQSPLAGESGTE
YITEVYSELPALPPLMPELDGEEVPEVPSVDPDIEVPRYFNKKDQSPLAGESGTE
MSGOTTPQVETEDTKPEVILMGSGSEVETPKDTQMGSGOTTPQVETEDTKPEVILM
GGSESEVETPKDTQMGSGOTTPQVETEDTKPEVILMGSGSEVETPKDTQMGSGFS
ETVIVEDTPKLVFHFDPNNEPKVEENREKPTKNITPLPATGDIENVLAFLITLIS
VLSIFSLKKNNKKV"

gene
CDS
5326..7596
/gene="cpa.12"
5326..7596
/note="putative collagen-binding protein"
/codon_start=1
/transl_table=11
/product="Cpa"
/protein_id="AAL86406.1"
/db_xref="GI:19224135"
/translation="MNNKLLQKODAPRVSNRKPQTLVGLVFLMFLVSSMRGA
QSIPEEKRIEESVPRIKSPDAPYGYDSYDSSHRYERFVADLNKLNLSKYS
YOVYCFNINSHYPRKNAKSPQWFKRVGDTGDTVAOTPKIRGESLNKLNLSIMN
AYPRNANGYMDKIEPLNALVTQOAVVYSDSGNFKTLMAEVLPSKDGKIDEBQVKM
REAVSKLISDDEETSKNKLPOGSKLNIIVPOKSYONILSAEVLPSPPAPQSPRP
PNOTKTSVIRKKAEDYSKLEGAFLRLTGEIDLPQEKVROSNKTEGKIELSNKT
YTLFESSPDGKIAEPIKFRVNNKVFIVOKDSOVENPKVEAPYSVEASVSDWD
SNYINPEFTPYGFEYAKNKKSSOVYCFNADLHSPSEDEGGTIDDISIMKCV
KYHTAGSDLEFKALBRPDNPEDFLHKIKVIEKYNKKGDSNGLTQFRAQL
AIYFTSDTLKTLKTYNNGKGYHFPESMDKELATVKELINYAQNDSAPOLNIDFE
VPNNKXOSLIGTEVHPDDLVDYIRMEDKOEVPVPTHSITVKKTVYVGEIGDITKGO
PELEIKDKTGOPVNTLAKTNODLVANDGYSRLKAGDTIRIEGLTGTSTYLTKE
AKDITVVDNRKVOEASASENVTADEVFENRKDLVPPGFTTDDGTLYMLLIVP
FGLVWFEGRKKLND"

gene
CDS
7589..8110
/gene="lepa"
7589..8110
/note="putative signal peptidase I"
/codon_start=1
/transl_table=11
/product="lepa"
/protein_id="AAL86407.1"
/db_xref="GI:19224136"
/translation="MTNYLNRLNENPLFKAFIRLVKISITIGFLGYLLFOYVFGWIV
LNTQMSPAVSAGGVLYRLTDRHINDVVVYVEDNLLKGRILAAQAGDVSFTQEG
LNLNGHPKEKEVYLYTPHSSGPNFKVPTGYTILINDYREERLDSRYGALPINOI
KGRISTLLVRGK"

gene
CDS
8132..9160
/gene="eftSL.A12"
8132..9160
/gene="eftSL.A12"
/codon_start=1
/transl_table=11
/product="eftSL.A"
/protein_id="AAL86408.1"
/db_xref="GI:19224137"

gene
CDS
9176..9901
/gene="eftSL.B"
9176..9901
/gene="eftSL.B"
/codon_start=1
/transl_table=11
/product="eftSL.B"
/protein_id="AAL86409.1"
/db_xref="GI:19224138"
/translation="MMFTIVQVINKAIDTLILFCLVYVFLAGEGLMDSYHLVQADA
SNFKFKTAQOQPEFDLALNEDVIGMLNIPGTHIDYPLVQGTNLEYINKAVDGSV
AMSGSLDLDRNHDFTDDYSLYGHMAGNARMEIPKLRKDFEKHNKALIEYKE
RKLTIVIFACLTDAFDOLVFNPNAITNODQOLVDYISKRSKQKPYKLRHHTFE
VAPSTCFNFTDNRYIVGTIQE"

gene
CDS
9936..10505
/note="orf2"
9936..10505
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL86410.1"
/db_xref="GI:19224139"
/translation="MFSVVMILMLAFNQTIVYAKDSTVOTSISENVYERADSTPF
SIALESIDAMKTEETITAGSGKASFPLEFTVGTGYTVYVQPSQNKVYQADTVF
DVLVYVYDEDDGLVAKVISIRAGDEKSAITPKRLVPIPPROPNIPKPLPLAG
EVKSLGILSIVLGLVLLVYRKLRKL"

gene
CDS
12254..15739
/gene="msmRL"
12254..15739
/complement(10664..11869)
/gene="msmRL"
/note="AraC-type DNA-binding domain-containing protein"
/codon_start=1
/transl_table=11
/product="msmRL"
/protein_id="AAL86411.1"
/db_xref="GI:19224140"
/translation="MVFEGLKHQTLHSQLPISVNSQDALKIYVGGNDYLLCYQ
FLKHALPQAAQVITFEYGLFEEGFSFMFLPCHYIIAIGPPYPSLKNQVDEQLANFL
KSHSRKEKLELSMAVLPPEPINNVNLIIDADEFDQETTCOOTIHOQLSSKSS
MTADPDILHLKHSISMAVLPPEPINNVNLIIDADEFDQETTCOOTIHOQLSSKSS
SIALREKRLTYIYLRLLRSEFVENTDVAKKYSLYKYMALNEBASDLTKYLRIRC
AATIHSESLFNKSIDSRQMTYSVLTVDSHLSYKLKYSIDAKRLVDSHLSRYFK
KYSNVSLOHYILSPKLEAQLLRGLPGEVAKSLYFDTTHFHKIFKKYTGSISSKD
YLAKYRDNI"

gene
CDS
12254..15739
/gene="prf2.12"
12254..15739
/gene="prf2.12"
/note="fibronection-binding protein; pfBpI"
/codon_start=1
/transl_table=11
/product="protein f2"
/protein_id="AAL86412.1"
/db_xref="GI:19224141"
/translation="MTQKNSYKLSFLSLTGTGLGLLVFGLSGVSVGHAETRNGAN
KQSFPEIKKVDQNNKPLPGATFSLTSKDGKTSVQFTSNDKGIVAQNIQDPQTYLK
ETAPAGDYKTSKRTPLVYVYENGTYKLEVPYNGEILSKSGSDVSSLOLEMPKKS
VSKYKTEVSSGAADYRNAAVEFKMSFLKQKSETINPGDTFLOIDRLNPKGI
SODIKTIYDSANSPALIGYAHNQLITTPDYIAGDKVLSABLSIFLENKEVLE
ENTSIKFSSTICQETTKGTGNVNLIGNSETESNTITNGLSNVGSGISTEYTEGE
FVWYVYVNPRTNIPIYATNMLWFGARASNTSOLENDANTSSALGEIIOYVEPDEK
LPSYGVYDKTLRLDITAGLNGEQMTKRORIDFGNNIQNKAFLIKYVGTQDSGK
PLVYQSLASFSGASEYAAFTPVGWNVYFQNEIALSPKSGSGSGSEFTKPSITVANL
KRYAOLRFRKMSSTDNVPDLPAAFEALSSNSQKLEASSNTOGEVHFKDLTSGTYLY
ETKAPKGYOQVETKLAITYVDTTKRPAEBENVYQWSPSSSVKVEANKRVTYVNHKEITTF
SGKKIENDRPDORPAKIOVOLLQNGQKMPNQIOETKIDWSYIHKDLPKIDAKMOE
YKSYVEEVNPDGYKVSYLGNIDIFNRETEFEVQNNFLEFGNAIKQSGSKIIIDE

OY		2025	TTCACCTGTTCAAAGAAACAGATTTCTGAAGCCTAIPAAGTTTAAAGTTTAATTAGCCAGAAGT	2088
Db		7350	TTTAATACCTGTGAAAAGACTGTGAAGTTAAAGATTATAATTAATCAACCGTGTAACAAGATTG	7409
OY		2085	AGCAAAATGCTCAGATTCCAAAAAGAGAAATAACAGATATAGACACTGTGGTTTTGAANA	2144
Db		7410	TCAAGAAAGCTCAATACAGCAAGTGAAGATGTCACAGCACAGAACAAAGAGTCACTTTGGAAA	7465
OY		2145	TAAATPAAGAGCCTGTGTCTCTACAGCATTTGATCAAAAGATCAATGGCTATCTTAGCTTT	2204
Db		7470	CCGGTAAMAAATTTGTGCCACCACAACTGTGTTTATTAACGATGATGGTGAACCTATCTGTGGTT	7528
OY		2205	GATAGTATTCGCTGGTATCAGATTGGGGATCTGGGGAAATTCACACGATPAAGATPAAGAAA	2264
Db		7530	ATTATTTGCTTTCCCATTTGGTTTGTGTATGTGTGTCTTCTTGGTGTGTAAGAAGACTPAAAAA	7588
OY		2265	ACATGACT 2272	.
Db		7590	TGACTAAT 7597	

/product="unknown"
 /protein_id="AAC97151.1"
 /db_xref="GI:4033714"
 /translation="MKRTKMLFSVYMLLMIAENQYVLADSVYVISISYENLEEAR
 GTSPTFALESIDAMKTIETITLGGSSKASPSPLNTFTVQITRYRIQKSNKDIO
 ADPTVATLVTYVYDEGDTLAKAVISRAGGEESATIEPKRRLVPIPPADPIPT
 PPLAGEVSLTLGIISIVLGLVLLVYVKKRSR"
 complement(1394..3169)
 /note="similar to electron transfer flavorprotein 1 of
 methylophilus electron; ORF 3"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAC97152.1"
 /db_xref="GI:4033715"
 /translation="MKRRKLLATLAIATALGMSMSONIKAFAGYIDGSTLVYKKT
 FSSYVDVNLMPKADYSEKVEADDNAKRTDGDIDIKRGVLDGLNKTTRYSSNDIT
 TAKEKSVNEFANVKEPGGVARYTVAEVRNNGKGIITVDSQOOWLENTTRYVNRKEGGE
 VYVYIIEGGESEKKRVLEKNSFDITSLIKERQVGTNGEHRQLFSTLLITPNCEEE
 KQAVNNITGGGTGKRVVLGEESFTLKDQGSVTLSPVIGIEYLTFTEDYVKKDYKTS

[illegible]

```

gene      complement(3191..3712)
          /gene="lepa"
CDS       complement(3191..3712)
          /gene="lepa"
          /note="ORF 4; similar to Staphylococcus aureus signal
          peptidase 1"
          /codon_start=1

```

REFERENCE
AUTHORS
TITLE
JOURNAL
2
AUTHORS
TITLE
JOURNAL
Submitted (19-FEB-1996) RWTH Aachen, Institute of Medical

```

/translation="MTNTLNLRNINSELEKAFIRLVLTAKSIIGFGLYILFOVFQEWMI  

NNDSDPLASAGDGLVYLRADIRSNINVVYVEVDNLTAKIRIAAGAGDELFPODEGG  

LLINHPPEKEVPEYLYTPHSSDPNFPYVPGTYFILNDYRBERLDSRYCALPINOI  

KKISTILRVRGI"
complement(3705, .6180)
/genes="cpa"
complement(3705, .5933)
CDS
gene

```

FEATURES	Location/Qualifiers
source	1. .10826

```

/note="collagen-binding protein of group A Streptococci"
/codon_start=1
/transl_table=11

```

```
gene      <1. .630  
          /gene="msmr"  
CDS       <1. .630
```

GAEDSVPNROSSIDYDPWGYDYSYPKGYPDYSPPLKTYHNKVAIEGSKDYQAYCFNL
TKHPPSKDSVSRSQWYKKLEGTNENFTKLADKRIEDGLOQNLIRILNGYPNNG
IMKGIDPLNALIVTQNAIWTDSAQINPDESFTEARSGINDQGLMKRALKELID

```

repeat_region
697, .730
AQLLLKRGTPVGEVAKSLFYDITTHFHKIFKKYTGISSKDYLAKYRDN1"

```

```
-10_signal complement(5972. .5977)
/gene="cpa"
```

```

CDS
    /1pl_type=inverted
    complement(790..1377)

```

```
-35_signal . complement(599/.6102)
/gene="cpa"
```

```
/codon_start=1
/transl_table=11
```

```
-10_signal complement(6138..6148)
/gene="cpa"
```


/note="extended promoter 2"
complement(6161..6166)
/gene="cpa"
/note="promoter 2"
6231..7898
/gene="nra"
6231..6241
/gene="nra"
/note="extended promoter 2"
6297..6302
/gene="nra"
/note="promoter 1"
6315..6325
/gene="nra"
/note="extended promoter 1"
6363..7898
/gene="nra"
/note="negative regulator of group A Streptococci"
/codon_start=1
/transl_table=1
/product="Nra"
/protein_id="AAC97147.1"
/db_xref="GI:4028948"
/translation="MPYVKKKDSFLVETYLEOSIRDKSELVILFFKSPITFSVAK
OTGLTAOLKYCKEELDPFGNNIDITIKKQITCEVKKPKERYILQVDTSTILK
LVFFIKNGTSSOPLIKSKYFLSSSSAIRRESLILKLPFGILVSKNTIVGEYRI
RYLAMLXSKFGVITYPLDLNDQIYRFLSOSATNRTSPWLEPSPYNNMLALSM
KRQFAVISIPQRIEFLQKLEIYDCLTRSRQYENASFJTFESQGLDYFLIYIT
NNSFASLOMTPOHLETCCHIEKENDTRLEPLIKRLPOLNHSKODLIKAMTSSK
FLNLOHVEIPEFSLPYTGNSLKALKNIYOMLAOLPGKRLHLEKLOLCSH
IEOLIKROPALTVLISNFINAKLITLDTIPRESQGIHFSYELLRDIDYIPSL
KPOLYTHSRILPFVKNDLVKGYVAEFSDEKDPYIASIONLITQLKDKKYODFLNE
QLO"
8006..8016
/note="ORF 5; extended -10 signal"
8024..8857
/note="ORF 5"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAC97154.1"
/db_xref="GI:4033717"
/translation="MDKIISIAQSGAFRAYVLDSTFVALAOKHNLTSSRYALGR
TLIANOLIANOKGDSKIYKVIQDSGFCHISADTKGHYKGIQNTQVNDIKTEGR
EVLGPFMGHNYTTIDYGTGNPTISTPLITGEIDEDRYVLTSEQPSALGVN
LLENDKVKVAGGMOVLPEASEEELAREKRLQEMPALSHLSAHNDALIELAY
LDEPYKRLSEEPISFOCDCSREREALMTLPKADLOAMIDEDKGAIEVCOFCGTLIP
I"
8884..9861
/note="ORF 6; similar to Acospirillum brasiliense nifH"
gene"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAC97155.1"
/db_xref="GI:4033718"
/translation="MILNSSFRIQDVEIDHRYVLAPMAGVTSNAPRTIAKEGAGIV
VMEIISERGLVYNNKATLHMLHIDENHPISIQLEGGDAGLKAAPFIQNTADIV
DINNGCVNRVYVNEAGAKVLRDPDKIYHLYKEVTSLVDIPLTKMSTVADSIAVE
NALAASAGVSALAMGRTREOMYTGCDHEHTLAKATIKIPIENGQVRSYODAK
FMIEIGDVAVMIGRAMNNPYLFTQINHPETGOELPDLPEAKKLIDADHLKRLIN
LKGETTAVREFGCIAPHYLRGTAGAAKRAVSAEFLAEALFEYVR"
9881..10522
/note="similar to Lactobacillus acidophilus da/dg-kinase"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAC97156.1"
/db_xref="GI:4033719"
/translation="MLIVLAGTIGAGSSSLAALGELHGLDVEYEAVDNNPVDLYYO
DPRKATLQIYFLNKRKSIKEAYQADNNILDSISIDEDELFLANKKNQNVKTEID
IYELLANMLEELGMPKRRPDLIYIDVFQKMLERIEKRGSEFQVDNLSIQY
HOVHGEYPTWYEDYKVSPPKKIDQNSLDYFQNPQDLATVTKMIDTKLKEHLH"

repeat_region 10598..10681
/note="putative transcription terminator"
/rpt_type="inverted
complement(10782..>10826)
/gene="ssba"
Query Match 31.7%; Score 720.8; DB 1; Length 10826;
Best local similarity 61.6%; Pred. No. 8.7e-113;
Matches 1372; Conservative 0; Mismatches 787; Indels 69; Gaps 11;
gene
61 AAAACTCAAAACCATTCATCTGCTACCTTAGTGGAGCTTTTAAATGATCTGCTTG 120
11
5894 AACAAAGCAGCACAAGACGATCGATGATGAAAGATTTTGGACCTTGAGCTG 5835
121 GTAACCTTCATGTTGCTGCTAAGACGCTTTTGGTTAGTAATCCTGAGCCAAAC 180
11
5834 ATAGGAATAGTAGGTTTCTATCAGACGCTTCGAGCTGAAGAAATCAGTACCAA- 5776
181 GCATTAATCCAGTTCAAGTTCGGAATACAGATGATGATGATGATCTTATAGTA 240
11
5775 -----TAGACAAAGCTCAATTCAAGATTATCCGTGATGATGATGATGATGAT 5721
241 GGCATCCATATTTTAAACAGTTTAGAGTACACACGATTTAAGGTTAACTTGAAGA 300
11
5720 GCGTACCACGACTATAGTCCGTTAAAGACTTACCATTAATTTAAAGTAAATTTAGAGGA 5661
301 AGTAGAAGTTTCAAGTTTATGCTTAAATTTAAAGAAAGCATTTCTCGATCAGAT 360
11
5660 AGTAGAAGTTTCAAGTACTGCTTAAATTTAAAGAAAGCATTTCTCGATCAGAT 5601
361 AGTAGTGTAAAGTGTATTAAGAAAGATGATGATGATGATGATGATGATGATGAT 420
11
5600 AGTGTAGATCAGATGATTAAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 5541
421 GCGATGAGCCCTAATATACGAGATGAGTAAATACGAGTAAATACGAGTAAATAT 480
11
5540 GCATATTAACCAAGATTAAGAGAGGAGGAGTAAAGAAATATATTTAGAGATTCCTAT 5481
481 AATGACATCCACAAGATGCAATGATATTAAGAGGCTTGAACCTTGAATGCTATC 540
11
5480 AATGATATCTTAATTAATCGTAATGAGTAAAGGATGATGATGATGATGATGAT 5421
541 AGATTTCAACAAGGCGGCTATGATGATGATGATGATGATGATGATGATGATGAT 600
11
5420 TTATGACTCAAAATGCTATTGTA-----TACTGATTAAGTCAAAATTAATCCGAT 5367
601 GAAAGTTTAAAGAGGAGTCAAGAAATGATGATGATGATGATGATGATGATGATG 660
11
5366 GAAAGTTTAAAGAGGAGTCAAGAAATGATGATGATGATGATGATGATGATGATG 5307
5306 CGAAAGCTTTAAAGAGGAGTCAAGAAATGATGATGATGATGATGATGATGATG 5247
721 CCGATGATTTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
11
5246 CCATCAGGTTTGGTTTAAAGTATGTAATGTAATGTAATGTAATGTAATGTAATG 5202
781 GGATACCAAAATCTTTGAGTGGTGGTTTACTTCAATTAACCAACCACTGAGAGAC 840
11
5201 ---TTCAAATCTTTGAGTGGTGGTTTACTTCAATTAACCAACCACTGAGAGAC- 516
841 CCACCAATGCTTCAAACTCAACGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCA 900
11
5145 -----AGAGCTCCGCTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 5091
901 GGTGATTAATCTTAATGCTTGAAGGTCGAATGATGATGATGATGATGATGATGATG 960
5090 GGTG---ACTTAATCTTGAAGGTCGAATGATGATGATGATGATGATGATGATGATG 5034
961 AGTTTCAAGGAGAGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
11
5033 GGTTCACAGAAAGAGCTTCAAGATTAATGTTTAAAGGAACTGCAATTCACCAAT 4974

OY	1021	GGACCTTACTACTTAACGAACTGATGATCTCCAGCTGGTTATGATATCGCAGCCATC	1089
Dd	4973	GGGACTTATACCTTACAGAAACATCATCTCCAGATGATATTAATATTCGGAGCCGATT	4914
OY	1081	ACTTTTAAGGTGAAGCTGGCAAGTGATATCTAT--TATGTATGAAAAACAGTTGAA	1137
Dd	4913	AAGTTTACGTAGAGAAATAAAAAGTATTTATTCGTCCAAAAGANTGGTTCTCAAGTGGAA	4854
OY	1138	AATCCCATTAAGAGATAGTAGACCTTACTCAGTAGAAGCATATATATGATTT--GAA	1194
Dd	4853	AATCCAAAACAAAGATGACAGAGCCATACACAGTGAAGGATATAAGACTTTATGAT	4794
OY	1195	GAATTTACCGTTTAACTACACAAAACATATCAAAATTTTATATGCAAAAATATAAAT	1254
Dd	4793	GAAAGATGACTCTCGGGTTTATCTCCATACGGAATAATCTTATTCAGTACCAATATAGAT	4734
OY	1255	GGAGCTTACAGGTTGTCTATTCCTTTATGACAGATCTTAATATCTCCACCACTGCA	1314
Dd	4733	AAAAGTTTCACAGTGTCTACTGCTCAATCTGTATTAACCTACACACCTGCATCATAT	4674
OY	1315	GATGGTGGAAAAACATGACTCCAGA--CTTACACAGAGAAGATTAATATACATCAT	1371
Dd	4673	GATATGGTGAACATATAATCCAGATCTAGTAGATGAAAAGATCAAGTACACACAT	4614
OY	1372	ATTGCAGTGTGACCTCTTTAAATATCTGTAACCAAGAGATCCGATCTGCACAT	1431
Dd	4613	ACGGAGGTACTGACTTGTTTAAATATCCGTAAAGACCAGATATCAAAATCCAGAAATC	4554
OY	1432	TTCTTTAAACATATCAAAAAGATATGAGAAGGTTACAGGAAAAAGCAAGCTAT	1491
Dd	4553	TTCTTTAAAGCATTTAAAAAGTATTTAAAAAGGCTATCAAGAAAAAGG-----TGAT	4500
OY	1492	GAGTATGATGCTTACTGACAGACAAATTCGCGCGCTACTCAGTTAGCAATATATAT	1551
Dd	4499	AGCTATATGATTTACAGAAACACAGTTTCGCGCGCTACTCAGCTGTATCATATAT	4440
OY	1552	TTCACTGATAGTGGCAATTAGA-----TAGGATTAACCTAAAGACTATCAT	1599
Dd	4439	TTTACAGCAGTGTGACTTTAAAACTTTAAAACTTTATTAACAAATGGAAAGTTACAT	4380
OY	1600	GGTTTGGAGACATGATGATAGTACTTTAGCACTGCTTAATATCCCTGTGAATACGT	1659
Dd	4379	GGATTTGATCATGATGATGAAAAAACCTACTGTCACAAAAAGATTAATATCTTATGCT	4320
OY	1660	CAAGTATATATCTCTCCACAGTAACTGATCTGATTTCTTTATTCGAAATACAAATAA	1719
Dd	4319	CAAAATGGCAATGCCCTCAACATCAAAATCTTGTTCTTCGTGACCTAAATTAAGCAAA	4260
OY	1720	TATCAATCTTATTTGAACTCAGTGGCATCCAGAAGTATAGTTGATATATTCGATG	1779
Dd	4259	GACCAATCTTATTTGGGACAGAAATGCCATCCAGATGATTTGGTTGACGTATTTCCGATG	4200
OY	1780	GAAATAAAA--AAGAGTTATACCTGTAACGTATAATTTAATGATGAGAAAAAGGTG	1836
Dd	4199	GAAATATAAAAAGCAAGATTTATTCAGTACTCACAGTTTGACAGTAAAAAAACAGTA	4140
OY	1837	ACGTGTTAGCTGGTGACAGAACTAAGATTTCCATTTTGAATTTGAATTTAAAAATAT	1896
Dd	4139	CTCGGTGAGTTGGGAGATTAACATTAAGGCTTCCAAATTTGAATTTGAATTTAAGATTA	4080
OY	1897	AAGCAGAAATTCCTTCTCAAACTGTTTAAACACATTAACAAACCTTCGATTTAAAGAT	1956
Dd	4079	ACTGGACAGCTATTTGTTAAACACTTAAAACTATAATCAAGATTTAGTAGCTTAAGAT	4020
OY	1957	GGTAAAGCAACATTAATTTAAACATGGGAAAGTTTACACTTCAAGTTTACAGAA	2016
Dd	4019	GGGAAATTTTATTATTAATCTTAACACATGGTGACACCATATAGAAATGAAGGTTTCCAGC	3960
OY	2017	GGTATTCCTTACCTTGCAAGAAACAGATTTCTGAAGCTTTAAGGTTAAAGTAAATAGC	2076
Dd	3959	GGATATTCCTTATACTCTGAAAGAGGCTGAAGCTATTAAGTATTAAGTACCGTTTATAC	3900

QY	2077	CAAGAAGTAGCAAAATGCTACAGTTCATCAAAAACGGAATATACAGATGAGCACTTCT	2133
Db	3899	AAAGTACTCAAGAAAGCCGACGTACGTAGGTAGGATATACGAAGAACAATAACTACT	3844
QY	2137	TTTGGAAATATATTAAGAGCGCTGTTGTTCTACAGAGAGTTGATCAAAAGATCAATGCTAT	2196
Db	3839	TTTGAAAACCGAAAAGATCTTGTCGCCACCACTGGTTTGACACAGATGGGGCTATCAT	3780
QY	2197	CTACGTTTGATATGATTAATCGCTGGTATATCAAGTTGGGGATCTGGGGAATTACACGATTAAG	2255
Db	3779	CTTTGGCTGTATATTAATCACTGTTCCACTGTTGGTTATATGTTGGCTATTTGGTGTAAAGG	3720
QY	2257	ATAAGAA	2264
Db	3719	TTAAAAA	3712
RESULT 5			
AE014138			
LOCUS	AE014138	52900 bp	DNA
DEFINITION	AE014138	AE014074	linear
ACCESSION	AE014138	AE014074	Genome.
VERSION	AE014138.1	GI:21903795	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
SOURCE			
gene			
CDS			

Query Match 29.2%; Score 664.4; DB 1; Length 52900;
Best Local Similarity 60.0%; Pred. No. 2,1e-103;
Matches 1318; Conservative 0; Mismatches 816; Indels 64; Gaps 10;

```
QY 61 AAAAAGTCAAAAGCAATTTACTGTCACCTTTAGTGGAGCTCTTTTAAATGATCTGCTTTG 120
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 15603 AACAAGACGACCAAGACATCGATTTACTGAAATATTTTACGCTTTGATGCTG 15662
QY 121 GTAACTTCATGTTGGTGAAGACGTTTGTGTAGTAAATCCTGAGCCCAAC 180
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 15663 ATAGAAATAGAGGCTTTTCTATCAGACCTTGGAGCTGAAGAACATCTGTCGCAAA- 15721
QY 181 GCAATTAATCCAGATTCAGTTCCGAAATACAGATGATGATATGATCTTATGATA 240
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 15722 -----TAAACAAGCTCAGTTCAAGATTTATCCGTGATGCTATGATCTTATTTCAA 15776
QY 241 GGGCATCCATTTTAAACAGTTTAGAGTACACAGATTTTAAAGGTTAACTTGAAGA 300
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 15777 GGCTACCCAGACTATGCTGCTTAAAGACTTACCTATATTTAAAGTAAATTTTGAATGA 15836
QY 301 AGTAGAAGTTATCAAGTTTATGCTTAAATTAAGAAGCAATTTCCCTGAGATCGAT 360
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 15837 AGTAGAGATTCAGACATCTGCTTTATTTTAAACAACCTTCCATCCAGCTCAGAT 15896
QY 361 AGTAGTGTAAAGAGGTATTAAGAAACATGATGAATCTCTACAAATTTGAAGATTAT 420
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 15897 AGTGTAGATCACAATGTTTAAAGAACTTAAAGAACTATGAAGAACTTTATACAGTTA 15956
QY 421 GCGATGACCCCTAGATTTAGCGAGATGACTAAATACAGATTACGAGCTGTATGTAT 480
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 15957 GCAATTAACCCAAAGTAAGAGCGAGCTTACACAAATATATGAGATCTCTCAT 16016
QY 481 AATGACATCCAAAATGGCAATGATTAATGGAAGCTTGAACCTTGAATGATC 540
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 16017 AATGATATCTTATGATCGTAAATGGAATTAAGAGGATGATGCTTAAAGCTAT 16076
QY 541 AGATATACAGAGAGCGGTATGTTACTATCTGATATGCTTCTTAAATCCAGAT 600
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 16077 TTATGACTCAAAATGCTATTTGTTACTATGCTATGATGCTTCTTAAATTTCTGATAC 16136
QY 601 GAAAGTTTAAAGGAGTCAAGAAATGATGTTGTTAGTACTCTCAATTTCTTTGATG 660
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 16137 AAAGCTTTCAACAAAGAAAGAAACAGATCTTAAATTAATCACAGCACTCACTAATG 16196
QY 661 CGTCAAGCTTTGAAGCAAGATGATCCGAATTTGGCACTAAATGCCAAACAGAT 720
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 16197 CGTATGCGCTAAAGATTAATTAATCAAGAGAGTAGAAATTTACGAATCAGATT 16256
QY 721 CCGGATGATTTTCAGTAACTATTTTGAAGTCTGAGGACAAAGGAGATTAATATATAA 780
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 16257 CCGCAAAATTAATCAAGTAAATTTTCAATCT-----TCGATTAAG 16298
QY 781 GGATACCAAAATCTTTGAGTGGTGGTTTACTCTCTAACAACCACTCCAGAGAC 840
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 16299 ACTTTCCAAAATCTTTTAAAGTGTGAGTATGTTCCGATCTATCCCCCAAAACCGAGA- 16357
QY 841 CCAACAATGCTCCAAATCAACCTCAAGACCTCAGTACTATTAAGATATGATATA 900
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 16358 -----NAGGCTCCGGCTAAAGCTGAAGAAACATCACTCATTTACAAATAATATGCGAA 16412
QY 901 GGTGATTAATCTTAAATGCTTGAAGTGAACATTTAGCTTGAAGGAGTAACGTAAT 960
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 16413 GGTGACTACTCTTAACTTTAGAGGAGCAACCTTAAAGCTGTCTCAAAATGAAGAGAT 16472
QY 961 AGTTTAAAGCAGAGCTTTAGCAGTAAATGATTTGGAGAAAGAAATGAATATCAT 1020
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 16473 GGTTTACAGAGAAATTTTGTAGTAAAGCTCAGAGAAAGTGAAGTCTCAAT 16532
QY 1021 GGAAGCTATTAATTAATGATGAATTCAGAGCTGTTATATGATGAGAGCCATC 1080
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 16533 GGGAGCTTAAGCTTTAAGTGAAGCTTAAGCTCTCAAGAGATATGAGTTGCCAACCCATT 16592
```

```
QY 1081 ACTTTAAGTTGAAGCTGGCAAGTGT---ATACATATTTATGATGAAAAACAGATTGAA 1137
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 16593 ACTTTTAAAGTGCAGCTGAGAAAGTTTATATAAATAAAGAGGTCAATTTGTAGAG 16652
QY 1138 AATCCCAATTAAGATATGATGAGCTTTACTCAGTACAGATATTAATGATTTGAAGAA 1197
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 16653 AACCAAAATTAAGAAATGCGAGAGCCATATCTTCTAAGAGCTTTCAATGACTTTGAAGAG 16712
QY 1198 TTTAGCGTTTAAAC---TACCAAACTATGCAAAATTTTATATGCAAAAAATTAAT 1254
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 16713 ATTGATTTATTTATGATTTTAAATACTATGTAAGTTTAACTATGCAAAAAATACAT 16772
QY 1255 GGAAGTTCAAGAGTTGCTATATGCTTAAATGACAGATCTAAATCTCCAGCAGCTGAA 1314
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 16773 GGAACCTAATCAAGATGCTACTAGTTTCAATGCTGATTTTACATCCAGCAGCTCATAT 16832
QY 1315 GATGCTGGAAACAATGACTCCGAG---CTTTCAACAGAGAGAGTAAGTAATACATCAT 1371
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 16833 GATCAGGAGCAAAATTAATGATCCGATGTCAGTAAAGTAAGATAAAGTATACAT 16892
QY 1372 ATTGAGGTGCTGACCTCTTAAATATCTGTAAGCAAGATACGATCCGATCAGACT 1431
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 16893 GTTTCTGCTATGATTTTGTATTAATATGACACCCACCAAGAGCAAGATGCTGATTTG 16952
QY 1432 TTCTTAAACATATCAAAAAAGTAATGAGAGGTTTACAGGAAAAAGACAAAGCTATT 1491
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 16953 TTCTTAAACATATCAAAAAATTCCTGATTAAGGTTATTAAGAAAGAGGGGGA-----T 17006
QY 1492 GAGTATAGTGTCTAATCTGACACCAATTCGCTGCTCTACTGATTAAGCAATATATAT 1551
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 17007 AACTATTAAGACATTAATACGTGAAGCTCAGTTTAAAGCACCAACCAATTAATTAATAC 17066
QY 1552 TTACATGATGCTGTAATTAAGATTAAGATTAAGTAACTA-----TAAAGACTATCAT 1599
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 17067 TATACATATGCTGCTGTTTAAACACCTTAAGCAATATATGATTAACAAAGGTTATCAT 17126
QY 1600 GGTTTTGAAGACATGAATGATAGTACTTTAGCACTTCTTAAATCCTTTGATAGATACGT 1659
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 17127 GGTTTTATTAACAGTGAATGATGCAACATTTAGCTGATCTCATGATGATGATCAGCTATGCT 17186
QY 1660 CAGATATGTAATCCTCCACAGCTAATGACCTGATTTCTTTATTCGGAATTAACAAATAA 1719
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 17187 GAGATGTCACCTTCAACAAAGTACCAAAATCTGATATTTCTTCTTAAATACAGATGA 17246
QY 1720 TATCAATCTCTTATTTGAAGCTGAGTCCAGAAAGATTAATGATTAATTTATTCATG 1779
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 17247 TACCAGACATTAATGGAAGCAGTATCATCAAAATGAATGATGATGATTTCTATG 17306
QY 1780 GAAGATTAAGAAAG---AAGTTATACCTGATCATATTAATTAACATTGAGAAAAAGCGTG 1836
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 17307 GAAGATTAAGAGGCTCTATTTATTCATTAATCAACAAGTTAATCTTAAACGTTT 17366
QY 1837 ACTGTTTACTGCTGTGACAGACATTAAGATTTTCATTTTGAATTTGAATTAATAAAT 1896
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 17367 ACTGGAATATTTGGAGTAAGAAAGAAATTTAACTTTAAATATCAATTAATATCTCT 17426
QY 1897 AAGCAAGATTTGCTTCTCAAGCTTAAAGATTAAGCAAAACATCCGATTAATGAAT 1956
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 17427 GATGACACATTAATAGTGAACATATCCAGCAAACTCTGGAGACCTACAGATTACGAT 17486
QY 1957 GGTAAAGCAACATTAATTTAAACATAGGGAAGATTAACTTCAAGGTTTACAGAA 2016
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 17487 GGAAGATCTACCTTCAATTAAGAGATGAGAAATCATTTGATGTTGGGGGCTACCTTCA 17546
QY 2017 GGTATTTCTTACTTGTGCAAGAAAGAGATTTGAGAGCTATAAGTTAAAGTTAATAGC 2076
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 17547 GGTACTCTTATGAATAATTAAGAAAGCGGCTTCAATTAATGAGGATGATGATTAATGGA 17606
QY 2077 CAAGAGTACCAATGCTTACAGTTTCAAAACAGGAATTAACAAGTGAAGTACACTTCT 2136
  || || || || || || || || || || || || || || || || || || || || || || ||
Db 17607 AAAAATGACCAAGATGAGAAAGGACGAAAGCCCTCACTTAAAGAAATGAGACTGATGCT 17666
QY 2137 TTTGAAATTAATTAAGAGCCTGTTGTTCTTACAGAGATGATCAAAAAGATCAATGGCTAT 2196
```

```

Db      17667 TTTAAACCAAGAAAGACTGTGTCCACCACTCG-----TTTGACACAGATGGGCGTAT 17722
Oy      2197 CTAGCTTGATGATGATTCGCTGATACGATTTGGGAGAT 2234
Db      17723 CTATCTTTGGTTGTTATTAATTACTGTTCATTTGGGTTAT 17760

RESULT 6
AP005141 311600 bp DNA linear BCT 13-JUN-2003
LOCUS Streptococcus pyogenes SSI-1 DNA, complete genome, section 1/6.
DEFINITION AP005141 BA000034
ACCESSION AP005141.1 GI:28810157
VERSION
KEYWORDS
SOURCE Streptococcus pyogenes SSI-1
ORGANISM Streptococcus pyogenes SSI-1
          Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
          Streptococcus.
REFERENCE
AUTHORS Nakagawa, I., Kurokawa, K., Yamashita, A., Nakata, M., Tomiyasu, Y.,
          Okahashi, N., Kawabata, S., Yamazaki, K., Shida, T., Yasunaga, T.,
          Hayashi, H., Hattori, M. and Hamada, S.
TITLE Genome Sequence of an M3 Strain of Streptococcus pyogenes Reveals a
          Large-Scale Genomic Rearrangement in Invasive Strains and New
          Insights into Phage Evolution
JOURNAL Genome Res. 13 (6), 1042-1055 (2003)
MEDLINE 22683278
AUTHORS 2 (bases 1 to 311600)
          Yamashita, A., Nakagawa, I., Kurokawa, K., Nakata, M., Tomiyasu, Y.,
          Yamazaki, K., Okahashi, N., Kawabata, S., Yasunaga, T., Hattori, M.,
          Hayashi, H. and Hamada, S.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-2002) Ken Kurokawa, Osaka University, Genome
          Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871,
          Japan (E-mail: ken@gen-info.osaka-u.ac.jp,
          URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
          Fax: 81-6-6879-2047)
COMMENT genome project
          This clone was isolated from a patient presenting with toxic shock
          like syndrome.
FEATURES
source 1..311600
          /organism="Streptococcus pyogenes SSI-1"
          /mol_type="genomic DNA"
          /strain="SSI-1"
          /db_xref="taxon:193567"
gene 232..1587
          /gene="SpS0001"
          /note="similar to GB:AAE71535.1 (AF255728) percent
          identity 100 in 451 aa"
          /codon_start=1
          /evidence=not_experimental
          /transl_table=1
          /product="chromosomal initiator protein"
          /protein_id="BAC63096.1"
          /db_xref="GI:28810158"
          /translation="MTENEQIFMNRVLEAQSOLKQATYEFVHADRLKVKDHAT
          YLDQKELFWEKNLKDYLITAGFEVNAQISVDYFEDLMEIQNQRKINQKPOQAL
          NSLPTVSDNSKSFENFIOGDENRNVAASIVANPGTTYNDLPFTWGGPGKTH
          LNAIGSVILENPNARIKYITTAENFNFYIHRIQPMDELKKEFRNLDLLDDIO
          SLAKRTLSGTQDEEFNFMAHHNNKOIVLSDTDPHLNDELRLVTRRWGILTVNI
          TTPDEFTRVALINKIQEYNTIFPDITIEYIAGQSDNVRLGALKDISIVANFKOI
          DTITVDIAAEAIRAKQDGPMTVPIPEIEIOAGYGVKVKIKATKRONTVLAR
          QVAMFLAREMTDNLSPKIGKEFGGRDSTVLAHVAKIKNMISQESRIETETIKNKI
          K"
          1742..2878
          /gene="SpS0002"
          1742..2878
          /gene="SpS0002"
          /note="similar to GB:AAL96838.1 (AE009953) percent

```

```

          identity 99 in 378 aa"
          /codon_start=1
          /evidence=not_experimental
          /transl_table=1
          /product="putative DNA polymerase III beta subunit"
          /protein_id="BAC63097.1"
          /db_xref="GI:28810159"
          /translation="MQQFSIRKTLFIRNLRNATKRAISIKKNAIPILSSKIEVPTGYT
          LTGNSGOISTENTIPVSNENAGLLTSPGALLBASFINILSSLPDININVEIEQH
          QVLTLSKSEITLKDQVDQYPRLOEVSTENPLIKTLKLSIIAETAFASLOESRP
          ILTGVLVHILSNHKDFKAVAFDSHMSORLITLADTSAADFVYIPGKSLREFAVFTD
          IEIVVEFSPQILFRSEHISFYTRLLEGNYPDPTRLMTEFEFEVFNQSLHAE
          RAFLSNATONGTAKLRTITONHISAHNSPVGKVNEDLDIVSOGSDLTISFNPYTL
          IESLKAIKSERVTKIHFLSPVNPFTLTTRDEDEESTIQLITTPVRTN"
          2953..3150
          /gene="SpS0003"
          2953..3150
          /gene="SpS0003"
          /note="similar to GB:AAK33148.1 (AE006472) percent
          identity 100 in 65 aa"
          /codon_start=1
          /evidence=not_experimental
          /transl_table=1
          /product="conserved hypothetical protein"
          /protein_id="BAC63098.1"
          /db_xref="GI:28810160"
          /translation="MYQIGSFVEWKKRHHACVYKETGKKANQWKLVRGADIKIOCTNC
          OHVITMSRYDERKRLKRVLP"
          3480..4595
          /gene="SpS0004"
          3480..4595
          /gene="SpS0004"
          /note="similar to GB:AAK33149.1 (AE006472) percent
          identity 98 in 371 aa"
          /codon_start=1
          /evidence=not_experimental
          /transl_table=1
          /product="putative GMP-binding protein"
          /protein_id="BAC63099.1"
          /db_xref="GI:28810161"
          /translation="MALTAGIVGLPNYKSTLFNATRKAGAEANYPATIDPNVGY
          EVPPERLQKTELITPKKTYPTPEFDIAGIVGASGEGLGKFLANIREVAIYH
          VVRAFDENVMREGEDEAPVADIDITINLELILADLESINKRYAVREKMAFQDQ
          KESVAEFENYLOKIKPVLDEGKSARTIETDEAKVVGILFTTRKPVLYVANVEDQV
          ANPGCIDYKQIRDPATENAEVYVISAAREEILLEDDEKEEFLEAIGITGECYDK
          LTRAYHILGLGITFYAGEKVRAMTRRGIKAPQAGIITHSDERGFTRAVTMSYDD
          LMTYSSERKAVKEAGRLRECKEYVVDGDIIMEFRNV"
          4665..5234
          /gene="SpS0005"
          4665..5234
          /gene="SpS0005"
          /note="similar to GB:AAK33151.1 (AE006472) percent
          identity 99 in 189 aa"
          /codon_start=1
          /evidence=not_experimental
          /transl_table=1
          /product="putative peptidyl-trna hydrolase"
          /protein_id="BAC63100.1"
          /db_xref="GI:28810162"
          /translation="MVKMIVGLNPGSKYKTKNIGFMAIDNIVKMLDYFTDCKNF
          KAOIGSFINIEKCFVPTFPNNSGIAVALLTYNIDITDLIVYDDIDMEVSKL
          RLRSKSGFGNGIKSIIAHIGQEPNRIRKVICGRPLKGMVIVHWVGQFTEDNIAI
          SLTIDRYNAKFTLQENDPEKTMQKENG"
          5237..8740
          /gene="SpS0006"
          5237..8740
          /gene="SpS0006"
          /note="similar to GB:AAK33151.1 (AE006472) percent
          identity 99 in 1167 aa"
          /codon_start=1
          /evidence=not_experimental
          /transl_table=1
          /product="putative transcription-repair coupling factor"

```



```

Db      112894 ATGTGTAATTTATCGATTTTAAATACATATGTGTAAGTTTACTATGCAAAAAATCTAAT 112953
QY      1255 GGAAGTTACAGAGTTGCTCTATTTGCTTTAATGACATCTTAAATCTCCACGACTCTGAA 1314
Db      112954 GGAAGTTAATCAAGTGTCTACGTTCAATGCTGATTTTACACTCCAGCTGACATAT 113013
QY      1315 GATGTTGGGAAACAAATGACTCCAGA---CTTTACACAGGAGTAATTAATACACTAT 1371
Db      113014 GATCAGGAGCAAAATATGATTCCTGATGTCAAGTAAATGAAGATTAAGATATACAT 113073
QY      1372 ATTCAGAGTCTGCTGCTCTTAAATATATCTGTAAGCAAGAGATACCGATCTGACACT 1431
Db      113074 GTTCTGCTATGATTTGTTATTAATATGACAGCCACCAAGGACAAAGATCTATTTTC 113133
QY      1432 TTCTTAAACATATCAAAAAAGTAATGAGAAGGTTACAGGGGAAAAAGGACACCTATT 1491
Db      113134 TTCTTAAACATATCAAAAAATTTCTGATTAAGAGGTTATTAAGAAAGGGGGA-----T 113187
QY      1492 GAGTATAGTGGTCTACTGAGACACATTTGGTGGCTACTGACTGATTAATATATAT 1551
Db      113188 ACCTATTAAGACATTTACTGAAAGCTCAGTTTAGAGCAGCAACCAATTAAGTATTTATAC 113247
QY      1552 TTCACGTAGTGTCTGAATTAAGATTAAGATAAATA-----AAAGACTATGAT 1599
Db      113248 TATACAGATAGTGTGATTTAAACACACTTAAGACATTAATGATTAACAGGGTATCAT 113307
QY      1600 GGTTTTGAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1659
Db      113308 GGTTTTGATTAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 113367
QY      1660 CAAGTATGATTAATCCCTCCACAGCTACGCTGATGATGATGATGATGATGATGATGATGAT 1719
Db      113368 GAGGATGCTACCTTACCAATGCTCAAAATCTAGATTTCTTGTCTTAATGACGTAGA 113427
QY      1720 TATCATCTCTTATTTGAACTAGTGCATCCAGAAATTTAGTATTTATTTGCTATG 1779
Db      113428 TACACGACATTTATTTGAAACGACAGATATCATCAATGATGATGATGATGATGATGATGAT 113487
QY      1780 GAGATTAATAAAG--AAGTATATCTGTAATCTAATTTAATGATGAGAAAAAGGCTG 1836
Db      113488 GAGGATTAACAGGCTCTTATTTCTTATTTACTCTCAAGTATTAATTTCTTAACCTGTT 113547
QY      1837 ACTGTTTACGTTGATGACAGAACTAAAGATTTCCATTTTGAATGATTAATAAATAAT 1896
Db      113548 ACTGGAATTTGACATGATGAGAAAAAGAAATTTAATCTTGAATATCAATTTAAATCTTCT 113607
QY      1897 AAGCAAGATTTGCTTTCTCAACTGTTTAAACAGATTAACAAACCTCGAATTTAAGAT 1956
Db      113608 GATGACAAAGCTATTAAGTGAACATATCCGACAACTCTGAGAACTACACATTCACAT 113667
QY      1957 GGTAAAGCAACCATTTAATTTAAACATGAGGGAAGTTTAAACACTTCAAGGTTTACCAGAA 2016
Db      113668 GGAAGAAAGCTACCTTACATTTAAAGGATGAGAAATCATGATTTGAGGGGCTACCTTCA 113727
QY      2017 GGTATTTCTTACCTTGTCAAAAGAAACATTTCTGAAGCTATTAAGTTAAAGTTAAATGAC 2076
Db      113728 GGTACTCTTATGAATTTACGAAACGGGCTCTTCAATTTATGAGTAAAGTAAATGGA 113787
QY      2077 CAAGAATGAGCAAAATGCTACAGATTTCAAAAACAGAAATTAACAGATGAGACACTTCT 2136
Db      113788 AAAAATGACACAGATGAGAAAAAGCAGAAAGCTCTGATTAAGAAATGAGACCTTACT 113847
QY      2137 TTTGAAATTAATTAAGAGCCTGTTGTTCTTACAGAGAGTTGATCAAAAGATCAATGCTAT 2196
Db      113848 TTTGAAACCGAAAAAGATCTTGCCCAACATG6---TTTGACACAGATGGGCTAT 113903
QY      2197 CTAGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2234
Db      113904 CTATCTTTGTTGTTATTTACTGTTTCCATTTGGGGTTAT 113941

```

```

AEO09963
LOCUS      AEO09963      11420 bp      DNA      linear      BCT_03-APR-2002
DEFINITION Streptococcus pyogenes strain MGAS8232, section 11 of 173 of the
complete genome.
ACCESSION  AEO09963
VERSION    AEO09963.1
KEYWORDS   GI:19747394
SOURCE     Streptococcus pyogenes MGAS8232
ORGANISM   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus pyogenes MGAS8232
REFERENCE  1 (bases 1 to 11420)
AUTHORS   Smoot,J.C., Barbican,K.D., Van Gompel,J.J., Smoot,L.M.,
Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M.,
Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M.,
Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.G. and Musser,J.M.
Genome sequence and comparative microarray analysis of serotype M18
group A Streptococcus strains associated with acute rheumatic fever
outbreaks
Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4668-4673 (2002)
JOURNAL   2197593
MEDLINE   11917108
PUBMED    2
REFERENCE 2 (bases 1 to 11420)
AUTHORS   Smoot,J.C., Barbican,K.D., Van Gompel,J.J., Smoot,L.M.,
Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M.,
Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M.,
Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.G. and Musser,J.M.
Direct Submission
Submitted (31-JAN-2002) Laboratory of Human Bacterial
Pathogenesis/Rocky Mountain Laboratories/NIAID/NIH, 903 S. 4th St.,
Hamilton, MT 59840, USA
FEATURES
source
location/Qualifiers
1..11420
/organism="Streptococcus pyogenes MGAS8232"
/mol_type="genomic DNA"
/strain="MGAS8232"
/db_xref="taxon:186103"
107..502
/feature="ssb"
/feature="synonym: spym18_0120"
/feature="ssb"
/feature="best blastp match gbl|AAC97155.1|AE006276.5
(AE006276) single-strand binding protein [Lactococcus
lactis subsp. lactis]"
/codon_start=1
/transl_table=1
/product="putative single-strand binding protein"
/protein_id="AAL96931.1"
/db_xref="GI:19747395"
/translation="MYNKVIAIGRLVAKPELVKATDQKVARLSLAVNRFRKNSGER
EAGFISVAVGKLAETLVASVSKSLMSIDELRTRKVDKGOVHYVTEVLCOSPOL
ESRAQRARENNVTDLVDVLEEDTLPF"
complement(757..1398)
/gene="spym18_0121"
complement(757..1398)
/gene="spym18_0121"
/feature="spym18_0121"
/feature="best blastp match gbl|AAC97155.1| (U49397) unknown
[Streptococcus pyogenes]"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAL96932.1"
/db_xref="GI:19747396"
/translation="MLIVLAGTIGAGKSLAALGELGTDFVEAVDNNPVLDTYYO
DPKRYAFLOIYFLNKRKRSIKRAYOANNTLDRSIFEDLFLKNTKNGVTKTELD
IXOELANMLELBGMRKRPDLITIDVSDKMLERERKGRSEVDGNSPEQYT
HGVHEPTWEDYKIVSPKRIIDGNSDFVQNPDLATVLMIDIKELHLL"
complement(1418..2395)
/gene="spym18_0122"
complement(1418..2395)
/gene="spym18_0122"
/feature="spym18_0122"
/feature="best blastp match gbl|AAC97155.1| (U49397) unknown

```

[Streptococcus pyogenes]
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="NAL96933.1"
/db_xref="gi:19747397"
/translation="MTKLNSSFRIGDVEIPHRVLAIPAGVTSNAPFTIKRPFAGVYVMEISSEGLIKYNNKTKLMLHIDENHPMSIQPGDABSLKAAPLOTNRKADVIDIMNCPKRVKRYKNEAGAKKLRDPDKITHYKETSVALDILEYTKMTGRMDSLAVERNALAESGVSAALAHGKTRRQMTGTCTDHEHTLRVSKALITKIPFGNGVSDAKEMFIEIIGDAVMIGRAANNPNYLFETQIMHFPESTQELDPDFAKDLIADHKLRLINLKGEIIVAREFGILAPHYLRGTAGAAKRGAVSRAETLAEVAEIFETVR"
/complement(2382..3254)
/gene="spym18_0123"
complement(2382..3254)
/gene="spym18_0123"
/complement(2382..3254)
/gene="spym18_0123"
/complement(2382..3254)
/note="best blastp match sp|Q92B45|HSLQ_STRPY_33_KDA
CHAPERONIN (HEAT SHOCK PROTEIN 33 HOMOLOG) (HSP33)"
/codon_start=1
/transl_table=11
/product="putative HSP33"
/protein_id="NAL96933.1"
/db_xref="gi:19747398"
/translation="MDKLIKSIAGSAGFRAYVLDSTETVALAQRKNTLSSSTVALGR
TLIAOIMLAANOQKSDSKITVYIGDSSPTGLISVADKRGHKYGIQNTGVDIKITAG
EVLVQEPNGNHFVYIIIDYGTGNPYSTPLITGEIGEDFAYLYTEEQGPDAIGLVN
LDEUDKQVAGGVFVQVLPGASRETEIARYKRLQEMPAISHLASNNHDALIEAY
GDEPKRLSEEPISLPQCCSRREREALMLTKRADLQAMIDEDKGAITVQGFCTKQ
FNESTLEALINDKA"
m18c.feature
complement(3380..4000)
/note="hra.2 (spym18_0124): sequence confirmed upstream
mutation gene fragment; best blastp match gb|AAC97147.1|
(U49397) Nra (Streptococcus pyogenes)"
/complement(4262..4915)
/gene="hra.1"
/note="synonym: spym18_0125"
complement(4262..4915)
/gene="hra.1"
/note="best blastp match gb|AAC97147.1| (U49397) Nra
[Streptococcus pyogenes]"
/codon_start=1
/transl_table=11
/product="N-terminal fragment of Nra"
/protein_id="NAL96933.1"
/db_xref="gi:19747399"
/translation="MPYVKKKKSDPLFVETVLEQSIIRKSELVLLLFKSPYIIFSHVAK
QTGLVAOLAKYCKELDPFNNMLDITIKKRIICCEYKPYKREYLAQLDYSTLIKL
LVFFKNTSTSOPLIKFKSKYFELSSSAVYRESLILKILREFGRKSVKNTIGVEYRI
RYLIAMSKRGIVYIPRDLHNDQIITRFLSQSNTNLTSTWLEPSEFYNMLALS"
5345..6919
/gene="cbp"
/note="synonym: spym18_0126"
5345..6919
/gene="cbp"
/note="best blastp match gb|AAC97148.1| (U49397) Cpa
[Streptococcus pyogenes]"
/codon_start=1
/transl_table=11
/product="putative collagen binding protein"
/protein_id="NAL96933.1"
/db_xref="gi:19747400"
/translation="MOKDKRTNGVSNANRRKROTTIGLKYFLFPVALIGVGSIRAF
GAERDSTTKRTSVIRKYRACDYSKLLKGGATLLAAGIDSGFDESGFESTSGQIKQ
LSDGYIILTERKSPQGEIAPITFKYTAGVFTKGDQGVENQNEVAEPYSVAY
NDPDSQGEINRKTPTPYGKYRANANGTSQVYVCFVWDLSKSPDSIDKGETIDDPFN
EGKTEYHIIAGDLFSAANNPRASTNDELISQVKKVLEKGRDSTTYKNTISVER
AATQALAIYFSDYDLNLADYHGFAGLTFPALNATKEIYVAAVDRANLPIINISLDEY
VATKQAYIIFSDYDLNLADYHGFAGLTFPALNATKEIYVAAVDRANLPIINISLDEY
FEIHLKSDQCAISCTYPTNSGELTYVDKATPFLKQGESIYEGSPSGSYETTFEG
ASDIEVSYNGNADGKATKASVKEDETFTEFNKKDLVPPGLTTGAIYLMILLVYL
LGLVWMLIGRGGAND"
6912..7433
/gene="lepa"

[illegible]


```

Db. 5673 TAACTGAAACTAAATCACTCAAGATATGATGGCGAACCTATTCTTTAAAGTTA 5732
Qy 1094 AAGTCGCAAGGTATATCTA---TTATGATGGAAGAAACAGATTGAAATCCCATTAAG 1130
Db 5733 CAGCTGGGAAAGTTTATTAAGAGTAAAGGCTGAGTTTGTAGAGAACCAATAATTAAG 5792
Qy 1151 AGATAGTAGAGCCCTTACTAGTACATATGATATGATTTGAGAAATTTGACGTTTAA 1210
Db 5793 AGGTCCGAGAACCTACTCTTACAGCCTTAATGATTTGATGATGATGTTTATTA 5852
Qy 1211 CTACACAAAC-----TATGCAAAATTTTATATGCAAAATTAATAATGAACTT 1261
Db 5853 ATCTAAAAACCTTCATCATATGCAAGTTTACTATGCAAAATTAATAATGAACTA 5912
Qy 1262 CACAGGTGCTATGCTTTATATGCAATCTAAATCTCCACGAGCTCTGAATGCTG 1321
Db 5913 GTCAAGTGTGTTACTGTTTCAACGTAACTTACATTCCTCCGACAGACAGTCTTACAAG 5972
Qy 1322 GGAAGAACATGACCTCCAGACTTTTACAACAGS---AGAAGTAAATATACATATTTG 1378
Db 5973 GAGAAACGATTTGATCCAGACTTTATAGAGGGAAGAAATCAATATATCTACATATTGG 6032
Qy 1379 GTGCTGACCTCTTAAATATCTGTGAACCCAGAGATACCGATCTGACACTTCTTAA 1438
Db 6033 GTGCTGACTTGTATAGTATGCTAATTAACCTAGGCGCTCAACAAGATGATTTATAT 6092
Qy 1439 AACATATCAAAAAGTAATGAGAGGTTACAGGAAAAAGCAACATTTAGATATA 1498
Db 6093 CTCAAGTAAAAAAAGTTTAAAGAAAGGTTATCTGTGATGATTCACAGCATATGCGAA-- 6150
Qy 1499 GTGCTTACTGAGACCAATTTGGTGGCTACTGCAATTTATATTTCACTG 1558
Db 6151 -----TTTAACTTCGGTGAATTTAGAGAGCTACCCATTTGCTATTTATTTATAG 6206
Qy 1559 ATAGTGTGTAATTAAGATTAAGATAAAGATATCATATGTTTGGAGACATGAATG 1618
Db 6207 ATAGTG-----TAGACCTGATATTTAGACAGATATCATATGTTTGGAGCTTTGACAA 6260
Qy 1619 ATAGTACTTTAGCAGTTGCTAAATCTTGTAGATACGCTCAAGATAG---TAATCTC 1675
Db 6261 CAGAGGCTTTAAATGCAACAAAGAAATAGTGTGCTTATGCTGAAGACAGGCTTAATCTTC 6320
Qy 1676 CACAGTAACTGACCTGATTTCTTATTCGAAATTAATCAATATCAATCTTATAG 1735
Db 6321 CTAAATTTCTAATTTGATTTTATGCTCTAATAGTAAATAGATATCAGCTTTAATG 6380
Qy 1736 GAACCTAGTGCATCCAGAGATTTAGTTGATTTATTCATGATGAAGATTAAGAA--- 1792
Db 6381 GAACACAATATCATCCAGAGCTTTAGTGATATTTATTCGAATGGAAGATTAACAGGCTC 6440
Qy 1793 AAGTTATACCTGTAATCAATTTAATTAACATGAGAAAAAGCGTACGCTTACCTGCTG 1852
Db 6441 CTATTATTCCTATTACTACAGAGTAACTATTTCTTAAACAGTGTGAACTATTTGAG 6500
Qy 1853 ACAGACTTAAAGATTTCATTTGAATTTGAATTTAAATAATTAAGCAAGATTTGCTT 1912
Db 6501 ATAGAAAAAAGATTTTAACTTTGAATTTAATTAATTTCTTGATGAGACACATTA 6560
Qy 1913 CTCAAACTGTTTAAACAGATTAACAAACCTCGAATTTTAAAGATGTAAGACCACTTA 1972
Db 6561 GTGAGACATATCCAGAACTCTGAGACATCAGAGTACAGATGGAAGAAAGCTACCTTCA 6620
Qy 1973 ATTTAAACATGGGAAAGTTTAACTTCAAGCTTTACGAAAGGTTATTTCTTACCTTG 2032
Db 6621 CATTTAAGATGGAGATCATATGATTTGAGGGGCTACCTTCAGTTACTCTATGAAA 6680
Qy 2033 TCAAGAAACAGATTTCTGAAGGCTATTAAGGTTAAATTAAGCAAGATAGCAATG 2092
Db 6681 TTACCGAAACGGGTCTTCAATATGAGTAAGTGTATTAAGAAAAATCACCAGATG 6740
Qy 2093 CTACAGTTTCAAAAAAGGAATTAACAAGTATGAGACACTTGTCTTTGAAAAATATAAG 2152

```

```

Db 6741 GAAAGCAGCAAAAGCCTCACTTAAGAGAGCAAGAAAGATTAATTTGAAAAACCGTAAG 6800
Qy 2153 AGCGTGTGTCTTACACAGACTTATCAAAAGATCAATGGCTATCTAGCTTGTATAGTTA 2212
Db 6801 ATCTGTCCACCAACATCTGTTTACACACAGATGGGGCTATCTATCTTTGTTTATTAAC 6860
Qy 2213 TCGCTGATCATGTTTGGGATCTGGGAAATTCACAGATTAAGATTAAGAA 2264
Db 6861 TTGTTCTACTGCGGTTATAGGTTTGGCTAATTTGGTCGTAAGAGCACTAAAAA 6912

RESULT 8
STREPINF 2845 bp DNA linear BCT 24-NOV-1993
LOCUS streptococcus pyogenes fibroneclin-binding protein (prtF) gene,
DEFINITION complete cds.
ACCESSION L10919.1 GI:425479
VERSION L10919
KEYWORDS adhesin; fibroneclin-binding protein; protein F; prtF gene.
SOURCE Streptococcus pyogenes
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 2845)
AUTHORS Sela,S., Aviv,A., Tovi,A., Burshtein,I., Caparon,M.G. and Hanski,E.
TITLE protein F: an adhesin of Streptococcus pyogenes binds fibroneclin
via two distinct domains
JOURNAL Mol. Microbiol. 10 (5), 1049-1055 (1993)
MEDLINE 95020565
PUBMED 7934855
COMMENT Original
FEATURES
source text: Streptococcus pyogenes (strain JRS75) DNA.
location/qualifiers
1..2845
/organism="Streptococcus pyogenes"
/mol_type="genomic DNA"
/strain="JRS75"
/db_xref="taxon:1314"
425..2404
/gene="prtF"
425..2404
/gene="prtF"
/function="binds fibroneclin"
/codon_start=1
/transl_table=11
/product="protein F"
/protein_id="AA26964.1"
/db_xref="GI:425480"
/translation="MEVREGEMNNKIFLNKEASFLAHTKKRRRPAVTLVGFYMLLAC
AGATGFOVAAYAADEKTVPSHSSNPPEPWGYDAYKEYGVYIMIRYHDLRVNLG
SRQVYVCNIOSSVPSOKNSFKMKMKKIDGNKSPDYATATTKLKEBELQRLSL
LYNAYPDNANGYKRGLEHLNATITVOTAVMHTSDNSOYOFETLWESAKBEKISRQV
TLMEALAKLIDPNLEATAVANKIPSGYRLNIFESENEAYOULLSAEYVDPDPPIGSE
SEHNPKPELDGTPIDPDPKHPDNLBPTLPVWLDEVEPEVSESLBPLPLMBE
LDGCEVEKPEIDLPIDVPRYEFNNKQDSPLAGSGSTEVITEVYGNOQNFVDIDKL
PNETGSGNVEETEDTKPEVLMGSGSEVEFTKDQTMGSGOTTPOVEDTDEPEVY
LMGSGSESVETKDTQTMGSGOTTPOIETEDTKPEVLMGSGSESVETKDTQTMGSG
OTTPOVEDTDEPEVYLMGSGSESVETKDTQTMGSGOTTPOIETEDTKPEVYLMGSG
SESEVETKDTQTMGSGSESEVETKDTQTMGSGOTTPOIETEDTKPEVYLMGSG
ATGDIENYLAFLFGLILISLISFLNKNKSV"
repeat_unit
1327..1457
/gene="prtF"
repeat_repeat
1631..2170
/feature="repeat domain 1"
repeat_unit
1631..2170
/feature="repeat domain 2"
BASE COUNT 977 a 472 c 616 g 780 t
ORIGIN
Query Match 9.58: Score 215.8; DB 1: Length 2845;
Best Local Similarity 57.28; Pred. No. 5.1e-27;
Matches 466; Conservative 0; Mismatches 322; Indels 27; Gaps 3;
Qy 51 GGTATTAAGTAAAACTCAAAACGATTTACTGTGCTTACTGAGAGCTCTTTTAATGAT 110

```

Db	487	GGCACACACAAAAAAGAAAGGCGCATTTGCTGTCACTTTAGTGGAGCTTTTTTAAGCT	546
OY	111	CTTGCGTTTGGTAACTTCACATGTTGGTGTCAAGACTGTTTTGGTTAGTAGAATCCTC	170
Db	547	TTTGGCATGCGCGGGTGTATCGGTTTTGGTCAAGTACCTTAATGCTGCCATGAGAAAGAC	606
OY	171	GAGCGCAACCCATAAATCCAGATTCAGTTCCGATATCAAGATGTAATGATATGAATC	230
Db	607	TGTGGCGGTCATAGTAGTCCAAATCCCT-----GAGTTTCCTTGGTCGGTTATGATGC	660
OY	211	TTATGTAGAGGGGATCCATATTATTAACAGCTTTAGAGTACACACGTTTAAGGTTAA	290
Db	661	TTATGTAGAGGAATATCCTGATATAAATTTTGGACTTCGTTATCATGCTTGAAGATAA	720
OY	291	CTTAGAGAGATAGAAATTTACAGTTTATTCGCTTAAATTAAAGAAAGCATTTCCCTC	350
Db	721	TTTAAATGGAGTGGAGACTACCAAGTTATTTGTTTAATATTCATCTAATTTATCCTAG	780
OY	351	CGATCAGATAGTAGTGTAAAAAGTGGTATAAAAAACATGATGGAATCTCTACAAATTT	410
Db	781	CCAAAAAATAGTTTCATTAATAAAATTTGGTTAAAAAAAATTAGGGAATAATGCCAATCCTT	840
OY	411	TGAAGATTATGCCAATGACCCCTAGCAATTACGGGAGATGAGCTAATACGAAGTTACGAGC	470
Db	841	TGTGAAATATCTCACACTACTAAACTTAGTAAAGAAAGTTGGAGCGAGAGCTCTTAAG	900
OY	471	TGTTATGTATATGAGCATCCACAAAATGCCAATGGTTATTATGGAAGGCTTGGAAACCCCTT	530
Db	901	TTTACTGTATATGCGTATCCCAACGATGCTAATAGTTATATGAAAGGACTTGGAGCATTT	960
OY	531	GAAATGCTATTCAGAGTTTACACAAGAGCGGGTATGGTACTATTTGTATAATAGCTCTTAATTC	590
Db	961	AAATGCTATTTACTGTATACCAATATATGCAAGTTGGCATACTACATATAATTCACAATATCA	1020
OY	591	TAAATCCAGATGAAGSTTTTAAAGGAGATCAGAAAGTACTTGGTACTTCTCAATTC	650
Db	1021	ATTTGAGCATTTATGGGAGTCTGAGGCGCAAGAAAGTTAA---GATTAATAGTCAACAGT	1077
OY	651	ATCTTGTATGGGCTCAAGCTTTGAAGCACTGATTGATCCGAATTTGGCAATTAATAAGCC	710
Db	1078	TACTTTAATGCGGTGAGGCATTGAAAAAACTAATTGATCCCAATTTAGAGCTACTGCAGT	1137
OY	711	AAAAACAAGTCCGAGTGAATTTTTCAGCTAAGTATTTTGGAGTGTAGAGCAAGGAGATAA	770
Db	1138	TAAATTAATCCCATCAGGCTATCGTTTAATATCTTTGAGTCTG-----	1181
OY	771	ATATATATTAAGGATATCCAAAATCTTTTGAAGTGGTGGTTTAACTCTTAACCAACAC	830
Db	1182	--AAAATGAACTTACCAAAAATCTTTTAAAGTCAATATATGACTCTGATGATGCCCTTAA	1239
OY	831	TTCAGGAGACCCACCAATGCTCCAAATCAACCTTC	865
Db	1240	ACGTGTGTAGACGTCAAGACATTAATTCCTTAATAATCTC	1274

RESULT 9			
SPSPB			
LOCUS			
DEFINITION	SPSPB	2263 bp	DNA
ACCESSION	X. pyogenes sfb gene for fibronectin-binding protein (partial).		
VERSION	K67947.1	GI:511149	
KEYWORDS	fibronectin-binding protein; sfb gene.		
SOURCE	Streptococcus pyogenes		
ORGANISM	Streptococcus pyogenes		
	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		
	Streptococcus.		
REFERENCE	1 (bases 1 to 2263)		
AUTHORS	Talay, S.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (31-JUL-1992)	S.R. Talay, GBF, Mascheroder Weg 1, 3300	
	Braunschweig, FRG		
REMARK	Revised by [4]		

[illegible]

Db 339 CTACAAGGACCTTACACAGATAT-----CATTAATTACAGCTAA 380
 QY 291 CTTAGAGAGAGTACAGTATTCAGTTTATTCCTTATTTAAAGAACATTTCTCT 350
 Db 381 TCTCAACGGAAGTAAACATATCAGCTTATGTTTATTTAAAGATTTGAACCTAA 440
 QY 351 CGGATCAGATAGTACTGTAAATAAGTGTATTTAAACATGATGTAATCTCACAANT 410
 Db 441 GAAACAGGATAGTATTTTTCGCAATTTGATTAAGAGTGAGTGTAGTGAAGACAT 500
 QY 411 TGAAGATTATCGATGAGCCCTAGAA-----TTACGGAGATGA 449
 Db 501 TGTAAATATGCTGATTAACCCGAAAGATATGAAGTTCAGAGATTATAGATGTA 560
 QY 450 GCTAAATCAGAGATTACAGCTGTTATGTATTAAGTGAATCCACAAATGCCATGTAT 509
 Db 561 ATTGAAAAAATATATTAAGAGTCTATATATGATGATCCGATTAACGGTAATGAT 620
 QY 510 TATGAGGCTTGGAGCCCTGAATGCTATGATTAACAGATTAACAGAGCGGCTATG 569
 Db 621 CATGAGGCTTATGAGCAATTAATGCTATCTTATTAACCTCAAAATGCTTTGTTACT 680
 QY 570 TTCTGATATGCTCTTATTTCTATTCAGATGAAGTTTTAAAGGAGTCAAGATGA 629
 Db 681 TTCTGATATGCTCTTATTTCTATTCAGATGAAGTTTTAAAGGAGTCAAGATGA 740
 QY 630 CTGCTGATGCTCTTATTTCTATTCAGATGAAGTTTTAAAGGAGTCAAGATGA 689
 Db 741 TAAATTAATGCTCTTATTTCTATTCAGATGAAGTTTTAAAGGAGTCAAGATGA 800
 QY 690 GAATTTGGCACTAAATGCGCAAAACAGTCCGATGATTTTCAAGTAAATTTTCA 749
 Db 801 TAAATTAATGCTCTTATTTCTATTCAGATGAAGTTTTAAAGGAGTCAAGATGA 860
 QY 750 GTCTGAGCAAGAGGAGATTAATAT 774
 Db 861 ATCTGAGATTAAGTATATCAAAAT 885
 RESULT 10
 LOCUS GGU31115 2066 bp DNA linear BCT 27-AUG-1996
 DEFINITION Group G streptococcus fibronectin binding protein (gfba) gene, complete cds.
 ACCESSION U31115 GI:950168
 VERSION U31115
 KEYWORDS
 ORGANISM
 SOURCE Streptococcus sp. 'group G'
 Streptococcus sp. 'group G'
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 REFERENCE 1 (bases 1 to 2066)
 AUTHORS Kline,J.B., Xu,S., Bisno,A.L. and Collins,C.M.
 TITLE Identification of a fibronectin-binding protein (gfba) in pathogenic group G streptococci
 JOURNAL Infect. Immun. 64 (6), 2122-2129 (1996)
 MEDLINE 96239026
 PUBMED 8675316
 REFERENCE 2 (bases 1 to 2066)
 AUTHORS Kline,J.B., Xu,S., Bisno,A.L. and Collins,C.M.
 TITLE Direct Submission
 JOURNAL Submitted (06-JUL-1995) James B. Kline, Microbiology & Immunology, University of Miami, 1600 NW 10th Ave, Miami, FL 33136, USA
 FEATURES
 source
 1. 2066
 /organism="Streptococcus sp. 'group G'"
 /mol_type="genomic DNA"
 /strain="1750"
 /db_xref="taxon:1320"
 /gene="gfba"
 /gene="gfba"
 252. 1994
 252. 1994
 CDS
 /gene="gfba"

/note="gfba"
 /codon_start=1
 /transl_table=11
 /product="group G streptococcal fibronectin binding protein"
 /protein_id="AAB06623.1"
 /db_xref="GI:950169"
 BASE COUNT 724 a 326 c 435 g 581 t
 ORIGIN

Query Match

Best Local Similarity 55.7%; Score 167.6; DB 1; Length 2066;
 Matches 398; Conservative 0; Mismatches 289; Indels 27; Gaps 3;

QY 61 AAAAACAAGATTTACTGCTACCTTAGTGGAGCTCTTTATGATCTTCCTTGG 120
 Db 321 AAAAAGAAAGCGATTTGCTGCTTACCTTAGTGGAGCTCTTTATGCTTTGGCATC 380
 QY 121 GTAACCTGATGCTGGGCTGTAAGCTGTTTGGTTTATGTAATCCCGAGCGCAAC 180
 Db 381 GCGGCTGCTATGCTTTGGTCAAGTACGCTATGCGCGATGACAGACGCTGCGGAT 440
 QY 181 GCATTAATCCAGATTTCAAGTTCGGAATACAGATGTAATGTAATCTTATGTA 240
 Db 441 TTTAAAGGCCAGATCA-----GATTATCCCTGGTANCGGTATAGATGCTTATCTG 494
 QY 241 GGGCATCCATATTATTAACAGTGTAGAGTACACACGATTTAAGGCTTACTTAAGA 300
 Db 495 GCGTTTAAAGATAT-----CATGATTTAAATGATTTTGAAGAT 536
 QY 301 AGTAGAAGTATCAAGTTTATGCTTATTTAAAGAAAGCAATTTCCCTCGATCAGAT 360
 Db 537 AGCAGCGCTTACCAAGTTTATGCTTATTTAGTAACAGAGAAACCGATAGGTTAT 596
 QY 361 AGTAGTGTAAAGAGTGTATTAAGAAACATGATGATCTTACAAATTTGAAGATTAT 420
 Db 597 GGTTCAGAAATTCGCTTCAAAAAGTGTATGATGATGATGATGATGATGATGAT 656
 QY 421 GCGATGAGCCCTAGAGTATGCGGAGATGAGTAAATCAGAGTATGAGCTGTATGAT 480
 Db 657 GCAGCTAATCCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 716
 QY 481 AATGACATCCCAAAATGCGCAATGATGATGATGATGATGATGATGATGATGATGAT 540
 Db 717 AATGATATCCCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 776
 QY 601 GAAAGTTTAAAGGAGTACAGAGTATGATGATGATGATGATGATGATGATGATGAT 660
 Db 834 GAAAGTGAAGTCTGAAATTAAGAGATGAGAGAAATTAAGAGTGAAGTGAAGT 893
 QY 661 GCTCAAGCTTGAACCACTGATGATGATGATGATGATGATGATGATGATGATGAT 720
 Db 894 GCTGATGCTCTTGAAGAAATTAATATCAGATGATGATGATGATGATGATGATGAT 953
 QY 721 CCGATGATTTTACCTAAGTATTTTGAAGTGTGAGCAAGGAGATTAATAT 774
 Db 954 CCGATGATTTAATGATTAATCTTTTATTTCAAAAGATTAATCTGATTAAT 1007

[illegible]

QY	231	TTATGTAAGAGGCGATCCATCTTTTAAACGTTTAGAGTAGACACACATTTAAAGGTTAA	290
Db	686	GTATAGCGCGAAGATGTTTAAAGTATCACAACTCAAGATA-----	726
QY	291	CTTAGAAGAGAGTAGAAGTTATTCAGATTTATGCTTAAATTTAAAGAACATTTCCCT	350
Db	727	-----ATGSAAGTAGAAGAGTATCAAGCGTATTTGTTTAACTGAAAAGTTTAGCCAAA	781
QY	351	CGATCAGATATGTAAGTGTAAAAAAGTGTAAAAAACATGATGAATCTTCACAAATT	410
Db	782	AAAAGAGAAAGTAGTCAGCTCTTAATTTGTTAAAAAAATTTAGATGGAGTACAGAACTT	841
QY	411	TGAAGATTATGCGATGAGCCCTCAATTTAAGCGGAGTAGAGCTAATATCAGAAGTTAGACG	470
Db	842	TAAAAAATACCTGAAAAACCTAGCTTTAGTGAGAGAGAAATTAAGACACATATATTGAA	901
QY	471	TGTTATGATATATGAGACATCCACAAAATGCCAAATGATTTATGGAAGCTTGGAAACCTT	530
Db	902	GGTTCTCTACAAATGATATCC---TAATATGATATGAATATATGAAGGGAATAGACCTCT	958
QY	531	GATGCTATCAGAGTTACACAGAGCGGATATGCTACTATTTGATATATGCTCTATTC	590
Db	959	CAGCGATATCTAGTATACACAAAATGCCATATGATATGATATGCTCGATATGTCACCGATTAA	1018
QY	591	TAATCCAGATTAAGTTTAAAGGAGTCCAGAAAGTCAAGTGGTGGTCTTCTCAAT	650
Db	1019	TGATATTTATATACCTCTTTACTAGTGAAGCCAAATGATCTTATATATACCACTAGCAGCT	1078
QY	651	ATCTTTGATGCGTCAAGCTTTGAAGCAACTGATTTGATCGAATTTGGCAACTAAATGCC	710
Db	1079	AACGTTAATGGGTAGACCTCTGAGAAAACATTAATTTCCCTGAT-----GAAAATTTGGT	1132
QY	711	AAACAGTTCCGAGTATTTTCAGCTATAGATTTTGGAGCTGAGAGCAAGAGATTA	770
Db	1133	TAAACAGTTCCATCGAATCTTAAAGTATGATTTTGGAGCT-----	1175
QY	771	ATATAATAAAGATACCAAAATCTTTTGAGTGGTGGTTAGTCCCTATAACCAACCAAC	830
Db	1176	-TCTGATATAGCTTACCAAAATCTTTTAAGTGCATATATGTACTATGATCTCTCTAA	1234
QY	831	TCACAGAGACCCACCAATGCTCCCAATTCACCTC	865
Db	1235	ACCTGATGATACGTACGACATATATCTTAAACATC	1269
RESULT 12			
AF009914			
LOCUS	AF009914	321 bp	DNA linear BCT 13-FEB-1998
DEFINITION	Streptococcus pyogenes fibrinectin binding protein F gene, partial cds.		
ACCESSION	AF009914	.	
VERSION	AF009914.1	GI:2267181	
KEYWORDS	Streptococcus pyogenes		
SOURCE	Streptococcus pyogenes		
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.		
REFERENCE	1 (bases 1 to 321)		
AUTHORS	Katerov,V., Andreev,A., Schalen,C. and Totolian,A.A.		
TITLE	Protein F, a fibrinectin-binding protein of Streptococcus pyogenes, also binds human fibrinogen: Isolation of the protein and mapping of the binding region		
JOURNAL	Microbiology 144 (Pt 1), 119-126 (1998)		
MEDLINE	98129085		
PubMed	9467904		
REFERENCE	2 (bases 1 to 321)		
AUTHORS	Katerov,V.E., Andreev,A.S., Schalen,C. and Totolian,A.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-JUN-1997) Medical Microbiology, Lund University, Solvegatan 23, Lund 22362, Sweden		
FEATURES	Location/Qualifiers		
SOURCE	1..321		

```

/organism="Streptococcus pyogenes
/mol_type="genomic DNA"
/strain="M4 strain Umea"
/db_xref="taxon:1314"
<1..>321
CDS

```

BASE COUNT	107 a	48 c	68 g	98
ORIGIN				

Query Match	4.38	Score	97	DB	1	Length	321
Best Local Similarity	60.48	Pred	NO	1.3e-06			
Matches	160	Conservative	0	Mismatches	105	Indels	0
						Gaps	0

OY		197	CAAGTTCGGATTAACGATGGTAGGTATGATCTTATGTAGAGCGCATCATTATATA	256
Db		57	CAGATCCGATTAATCCCCTGGTATGGTTATTATGCTTATGTATGGTGGATATCCTGGATATG	116
OY		257	AACAGTTAGAGTAGCACACGATTTAAGGGTTAACTTGAAAGAAGTAGAACATTACAAG	316
Db		117	ATATCAGTAACATTAATTCATCATGATTATTAAGGGTAATCTTAACGAGTCAGCTCAACAG	176
OY		317	TTTATGCGTTAAATTTTAAAGAAAGCATTTCCCTCTCGGATCAGATAGTAGTGTAAAAAGT	376
Db		177	TTTTATGCTTAAATTTTCAAAATAATTTTCCGTACAAATGTGAAGAAAGCCTTACTCAAAAAT	236
OY		377	GGTATAAAAAACATGATGGAATCTCTACAAAATTTGAAGATTTATGCAGTACGCCCTTAGNA	436
Db		237	GGTTTAAAAAAGCCGAGGGGAATTCGTATACTTTGGACCTTATGCATGATTCACAAGAG	296
OY		437	TTACGGGAGATGACGCTAAATCAGAA	461
Db		297	TTCAGAGGAGAAAGAACTTAGTCAAAA	321

RESULT	13
AF009913	.
LOCUS	
DEFINITION	Streptococcus pyogenes fibronectin binding protein F gene, partial
	313 bp DNA linear BCT 13-FBB-1991
AF009913	

ACCESSION	AF009913
VERSION	AF009913.1
KEYWORDS	GI:2267179
SOURCE	.
ORGANISM	Streptococcus pyogenes
	Streptococcus pyogenes
	Streptococcus pyogenes

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 313)	Katrev, V., Andreev, A., Schalen, C. and Totolian, A.A.	Protein F, a fibronectin-binding protein of <i>Streptococcus pyogenes</i>

JOURNAL
MEDLINE
PUBMED
98129085
Microbiology 144 (Pt 1), 119-126 (1998)
9467904
2 (bases 1 to 313)
Katerov,V.E., Andreev,A.S., Schalén,C. and Totollian,A.A.
REFERENCE
AUTHORS
TITLE
Direct Submission
Submitted (23-JUN-1997) Medical Microbiology, Lund University,
JOURNAL
Solvegatan 25, Lund 22362, Sweden

```

1. .313
   /organism="Streptococcus pyogenes"
   /mol_type="genomic DNA"
   /strain="M4 strain 281C"
   /db_xref="taxon:1314"
<1. .313
CDS

```

```

/codon_start=2
/transl_table=11
/product="Fibronectin binding protein F"
/protein_id="AAC38160.1"
/db_xref="GI:2267180"
/translation="FGGVAYSADERTVPEKSPDDPYMGYGDAYGKGYGDIKSKYY
HDLVNLNGSQVYVCFNFIQIKRPPVIVKSVYIQKMEKAGNSDTFELHAMNRYQGE
EL"
BASE COUNT      102 a      47 c      68 g      96 t
ORIGIN

```

Query Match	4.1%;	Score 92.8%;	DB 1;	Length 313;
Best Local Similarity	60.2%;	Pred. No. 6	6e-06;	
Matches 154;	Conservative 0;	Mismatches 102;	Indels 0;	Gaps 0;

QY	197	CAAGTTGGGAATACAGATGGTATGATATGATCTTATGTGAAGGGCATCCAAATTATA	256
Db	57	CAGATCCAGATTAATCCCTGGATGGTTATGATGCTTATGGTAAGGGTATCCGATATG	116
QY	257	AACAGTTTAGAGTACACACAGATTAAAGGTTAACTTAGAAGCAATGAAATTATCAAG	316
Db	117	ATATCACTAAGTATATATCATGATCTTAAAGGGTAATCAAAACGGAAGTCAGTCCACAG	176
QY	317	TTTATTCCTTAATTTAAAGAAACCATTTCTCCTCGSATCAGATAGTAGTTTAAAGT	376
Db	177	TTTATTCGCTTAATATTCACAAAAATTTTTCGTCATATGAAAAAGGCTTACTCAAAAAT	236
QY	377	GGTATAAAAAACATGATGGAATCTCTCAAAATTTTGAAGATTATGCATGAGCCCTGAA	436
Db	237	GGTTTAAAAAAGCCGAGGGGAATTCGTACTTCTTGGAACCTTACATGAAATCCAAAG	296
QY	437	TTACGGGAGATGAGCT	452
Db	297	TTCCAAGGAGAGAAGCT	312

RESULT 14	AP004172/c	LOCUS	293650 bp	DNA	Linear	BCT 10-DEC-2002
DEFINITION	AP004172	Mycoplasma penetrans genomic DNA, complete sequence, section 3/5.				
ACCESSION	AP004172	BA000026				
VERSION	AP004172.1	GI:26453922				

REFERENCE
1
ORGANISM SOURCE
Mycoplasma penetrans
Mycoplasma penetrans
Bacteria; Firmicutes; Molluscates; Mycoplasmataceae; Mycoplasma.

AUTHORS Sasaki, Y., Ishikawa, J., Yamashita, A., Oshima, K., Kenri, T., Furuya, K., Yoshino, C., Horino, A., Shibb, T., Sasaki, T. and Hatori, M.

TITLE The complete genomic sequence of *Mycoplasma penetrans*, an intracellular bacterial pathogen in humans

JOURNAL Nucleic Acids Res. 30 (23), 5293-5300 (2002)

PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

12466555
2 (bases 1 to 293650)
Sasaki, Y
Direct Submission
Submitted (18-SEP-2001) Yuko Sasaki, National Institute of

Infection Control: 4-7-1 Gakuen, Musashimurayama,
Japan (E-mail: yuko@nih.go.jp, Tel: 81-42-561-0771,
Fax: 81-42-563-3315) Tokyo 208-0011,

```

FEATURES
source
location/Qualifiers
1..293650
/organism="Mycoplasma penetrans"
/mol_type="genomic DNA"
/strain="HF-2"
/specific_host="Homo sapiens"
/db_xref="taxon:28227"
/note="clinical isolate"
104..1006
/gene="MPE4630"
gene

```

CDS
104. .1006
/gene="MYPE4630"
/note="paralog family"
/codon_start=1
/evidence=not_experimental
/transl_table=4
/product="hypothetical protein"
/protein_id="BAC44253.1"
/db_xref="GI:26453923"
/translation="MFKRRKNINDPLKNKYLSPKGEFNDNDLFDYSQMMQDLEERT
KEHOKVKKRLSLNVDLKLKEVEBSKDLKTTINDPCITPKDKKKKIYWSLFLSLI
SRLLHGFYLFALIAVYVSABHLRPVYGLMKRRLLYITTVVAVINPFLAALIF
SGEKKSKVGRSIIIGIATANAIYRETKLTKYLLDLYNDSNDADKENDPKIKKO
VILLINQYFHNLLDTPFLDKELTNDSDVNNLKASRSKSIINKNFICINLYNEYQRP
LHMFVSTFIYAGIAYLYFLR"
1197. .1454
/gene="MYPE4640"
1197. .1454
/gene="MYPE4640"
/codon_start=1
/evidence=not_experimental
/transl_table=4
/product="hypothetical protein"
/protein_id="BAC44254.1"
/db_xref="GI:26453924"
/translation="MFVNENKIKSTFOKEIKNCWEKMFNEFCNFYITSVKNNLIYF
AVHLKIANKISNCTDFVLEKRSIFRFKILIGDLNA"
1447. .2370
/gene="MYPE4650"
1447. .2370
/gene="MYPE4650"
/note="MYPE4630 paralog"
/codon_start=1
/evidence=not_experimental
/transl_table=4
/product="hypothetical protein"
/protein_id="BAC44255.1"
/db_xref="GI:26453925"
/translation="MKNPNDFPKTKNFKRRKTELNGEKKVTDGLGLIANGVLKO
FESKKNQKHLKFLSLNVDINKELQONNLSKTIWPCITPKNNKKNKLYWSLFI
ALISRFEGHILVPTLAVLSKYLPLSGSTTNHLSREDVYVSFVCLIGNLIF
VLFTSGRANKERYNRRTIISTSNIFLYRKRTIIVKCYWLNSNNDQMNEDKE
LIEETKHNILINGFVRLGLTFLNNTKLENGIKNDKRYKKSLYMENIVSYIYC
FYINSIALIYLFESIFIVAGFYLYQTFPO"
2480. .3340
/gene="MYPE4660"
2480. .3340
/gene="MYPE4660"
/note="MYPE4630 paralog"
/codon_start=1
/evidence=not_experimental
/transl_table=4
/product="hypothetical protein"
/protein_id="BAC44256.1"
/db_xref="GI:26453926"
/translation="MFVNENKSTFKKEEIKNYKDFDSRDSSEHLNORIEVOKIR
KSLISNVDFEKLNEIQOSKFEOTLNEPCIPIDKKNKRIYLSIFISILRPIYVHG
FMVLCTVLLFLNKLIPYKSEWYIILFPLAFYLNSTICYOLIFYNKHETKYI
RGLIGESTNIEFYIGOTRRLYKKSQOLMSNKEEDNKEFVPLINAVENYILRTF
FSNTEKPEKDINSSTOKTKIYKINFRIFYGLIYNTLYPYCLIVSMIFYICIGIIT
VOFLINILF"
3598. .4521
/gene="MYPE4680"
3598. .4521
/gene="MYPE4680"
/note="MYPE4630 paralog"
/codon_start=1
/evidence=not_experimental
/transl_table=4
/product="hypothetical protein"
/protein_id="BAC44257.1"
/db_xref="GI:26453927"
/translation="MLRKSSEKSKVFKTIKMKNSVDENTEVKNVDIJTFWDLLEK

gene
4647. .5936
/gene="MYPE4690"
4647. .5936
/gene="MYPE4690"
/standard_name="pgl"
/codon_start=1
/evidence=not_experimental
/transl_table=4
/product="glucose-6-phosphate isomerase"
/protein_id="BAC44258.1"
/db_xref="GI:26453928"
/translation="MKLNFNPLBGNKSTFKSYASKIETLNNISIEKKCKGASM
LGWMDPSKYNKNEIKELINVSQEWNNKIRNVVGLIGGSYIGVAGIDWLVPEFN
REKEIYVSSMSYVSLERLKEKEDFLIYISKSGTLEIGVAFRLFPYSLFEPFG
SEGAKERVVAITDKEKGLADQIDQTFESIPDPVGRFSATFPVGLFAMGVGLD
VKNVLRGCAKAIEDTRKEDPIENKTAQVAAALRHYMSKKEKNEYCVGLDLPFTE
HLKQLFASEGEKGLDPVNLFTFDLHLSVGOFLQEGNOIFEFNCISITKNPLNVI
NEFLKDEGLGFLNGRKLDYINKVASSTIDHHIDKIDITIQINLDNNRESFGILY
SMFSKVVAMSGLLKYNPDPQGVFAKSMFTKLK"
6048. .7133
/gene="MYPE4700"
6048. .7133
/gene="MYPE4700"
/codon_start=1
/evidence=not_experimental
/transl_table=4
/product="conserved hypothetical protein"
/protein_id="BAC44259.1"
/db_xref="GI:26453929"
/translation="MYSFNNKKEKKKNFPPREDFNGLNIAIVNLYIYAVLLIS
VPSIVIGIISFTPKIKAKTSFYLYAFSSAVFLMGTVGLIEGTEFTFHEIDNS
EGINOLIMAGIAGALAGLVAIVFRYFVFEVSGEHTHDSHDHIFMLSDPN
PKAAMLVILILSHRTIDGFLGVAKMSTGADNLGLFTIFLIHIEVEVLIYVRQ
VOYQGRKALYFVTLTAIVPMVIGAVINQYLAOVAMILPEVNSGSGSIARVQ
TELPEPLHRYKMSKDWYKLIICVSGIVFALFILSPTHHEVHEAHHDHETTO
YKLIANPKOYLSLNNFNIVANNQPV"
7204. .8640
/gene="MYPE4710"
7204. .8640
/gene="MYPE4710"
/standard_name="gap"
/codon_start=1
/evidence=not_experimental
/transl_table=4
/product="NADP-dependent glyceraldehyde-3-phosphate
dehydrogenase"
/protein_id="BAC44260.1"
/db_xref="GI:26453930"
/translation="MNTFYKKSATINCKFVNSEKTIPIYSTIDSTIGSVAALSKD
IMAFEGAHFLNMLNLPENRIKIKFEAFYIAEKELFATLMSYELGKSYNALK
EVERSEYENLEFIEVEKEFVNPVIGPEVNNIKGTGTFYVNPVGVVALAPFNPPI
NGLKAIITFLVNTGIVKPAQOSLVSQAKFEDANFAGVFNVTGSGSEIGD
YLENNRIGAGFAITGSSDIEGFLDALPKPKPYLVEGLADAIIVNMADVGAKEII
KCAPYSGORCAIRIVLYTEKRIADDLIHLIKENOKIKVGNPLNNDPTPLDKAV
DNPIDLIDAIHGANLYVGCNIEGYNLNRHILDNVNSKSLAMEEPPGRLPIIRV
STINERIALANSEYGLQGSVFTKDLDEARITAKYLDGTVINNGSSRGDPIDFPFIG
IKKSGFVGIGIRDSLKAMTRVGIENE"
complement(8788. .9687)
/gene="MYPE4720"
complement(8788. .9687)
/gene="MYPE4720"
/note="paralog family"
/codon_start=1
/evidence=not_experimental
/transl_table=4
/product="oxidoreductase-aldol/keto reductase family"
/protein_id="BAC44261.1"
/db_xref="GI:26453931"
OKTNNSHKLKFSISLNVNLSISKIETLNFENFWTLKEPCIPINNNKKNKIYWSLFI
ALISRFEGHILFALIAVLSIYKLYVNSSEELIYRRKMVHSIIAISAGLNFVSG
LYWYCKDKDTEFRRTIISTSNIFPQONTSEKVKCFILSTNDIKTYTNLKNFKY
NKKLFYIYVHHHNLIDTPEDFETNOSKNSVSKNRIIKPLKKEHVLISIKFY
SIYSTLIHMLISVLIYAGFYLYQVFLONL"
4647. .5936
/gene="MYPE4690"
4647. .5936
/gene="MYPE4690"
/standard_name="pgl"
/codon_start=1
/evidence=not_experimental
/transl_table=4
/product="glucose-6-phosphate isomerase"
/protein_id="BAC44258.1"
/db_xref="GI:26453928"
/translation="MKLNFNPLBGNKSTFKSYASKIETLNNISIEKKCKGASM
LGWMDPSKYNKNEIKELINVSQEWNNKIRNVVGLIGGSYIGVAGIDWLVPEFN
REKEIYVSSMSYVSLERLKEKEDFLIYISKSGTLEIGVAFRLFPYSLFEPFG
SEGAKERVVAITDKEKGLADQIDQTFESIPDPVGRFSATFPVGLFAMGVGLD
VKNVLRGCAKAIEDTRKEDPIENKTAQVAAALRHYMSKKEKNEYCVGLDLPFTE
HLKQLFASEGEKGLDPVNLFTFDLHLSVGOFLQEGNOIFEFNCISITKNPLNVI
NEFLKDEGLGFLNGRKLDYINKVASSTIDHHIDKIDITIQINLDNNRESFGILY
SMFSKVVAMSGLLKYNPDPQGVFAKSMFTKLK"

[illegible]

Dd	224812	TGAATCAGACAGAAAAATGAAAGATGCAATTTTATGTATGAGAAGATGATCCCTGGAAAACCTGA	224753
Oy	1764	TGATAATTAATTCGTATGAGAAATTAATAAAGAACCTTATACCTGTACTCATTAATTTAAcATT	1823
Dd	224752	AGATCGAAGAAATATCTCTGATGATGAAATCAGATGATTTAGATATCTCGATGAAGATGAAAA	224693
Oy	1824	GAGAAAAACGGTGACTGGTTTAGCTGCGTGACAGAACTTAAGAATTTCCATTGTTGAATTGA	1883
Dd	224692	TGAAGATTAATGATTTAGATGAAATCTGAGAGATGAACAATGAAATTAACGAAATTAAGATTTC	224633
Oy	1884	AFTTAAAAATPATATAGCAAGAAATTCCTTTCGAAACCTGTTAAACAGATTAACAAACCT	1943
Dd	224632	TGAAGATATATATTATAGAGAGACTANTGAGTCTGTATGAAATTTAGAAAGATTCTCACAGAGAACC	224573
Oy	1944	CGAATTTTAAAG---ATGGTAAAGCAACCACTTAATTTAAACATGGGAAAGTTTAAcACT	2000
Dd	224572	TGAAGATGAGAAATATTCCTGATGATTCAGAGAAAATAGTGAAGATGATTTCGAGAACT	224513
Oy	2001	TCAAAGTTTACCAGAAAGGTTATTCCTTACCTGTCCAAGAAACAGATTCTGAAAGCTATTA	2060
Dd	224512	TTATGCTAATCACTGAGATGAAAGATGGAATCAATCATTAATGAAGAAACGAAAGACTATGAAATC	224453
Oy	2061	GSTTAAATTTATATACCCAGAAAGTGAACAAATGCTCAGTTGCCAAAAAACAGAAATACAAcAG	2120
Dd	224452	TGATGATATAAATGATGAGAGATTACTTGATATATTCAGAGAAATTAAGAAATGAAATCTTA	224393
Oy	2121	TGATGAGACACTTGCTTTTGAAAAATATAAAGA	2153
Dd	224392	TGATGAGAAAGATGGTTCTGAGAAAGAACTGAAGA	224360
RESULT 15			
LOCUS	AL935272	145265 bp	DNA linear HTG 28-MAY-2003
DEFINITION	Danio rerio clone CH211-266K22, *** SEQUENCING IN PROGRESS ***, 3		
VERSION	AL935272		
KEYWORDS	HTG; HGS; PHASEI; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	Danio rerio (zebrafish)		
ORGANISM	Danio rerio		
REFERENCE			
AUTHORS	Andrew R.		
JOURNAL	Submitted (27-MAY-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Apr 5, 2003 this sequence version replaced gi:28949357.		
COMMENT			
	----- Genome Center -----		
	Center: Wellcome Trust Sanger Institute		
	Center code: SC		
	Web site: http://www.sanger.ac.uk		
	Contact: zfish-help@sanger.ac.uk		
	----- Project Information -----		
	Center project name: zc266K22		
	----- Summary Statistics -----		
	Assembly program: XCAP4; version 4.5		
	Chemistry: Dye-terminator; 100% of reads		
	Consensus quality: 144211 bases at least Q40		
	Consensus quality: 144460 bases at least Q30		
	Consensus quality: 144678 bases at least Q20		
	Insert size: 145065; sum-of-contigs		
	Insert size: 156629; 0.9% error; agarose-fp		
	Quality coverage: 8.29x in Q20 bases; sum-of-contigs Quality		
	coverage: 7.97x in Q20 bases; agarose-fp		
	----- NOTE: This is a 'working draft' sequence. It currently		
	* consists of 3 contigs. The true order of the pieces		
	* is not known and their order in this sequence record is		

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 97678: contig of 97678 bp in length
* 97679 97778: gap of 100 bp
* 97779 101129: contig of 3351 bp in length
* 101130 101229: gap of 100 bp
* 101230 145265: contig of 44036 bp in length.

FEATURES

source

1. 145265
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-266K22"
/clone_1lb="CHORI-211"

misc_feature

1. 97678
/note="assembly_fragment:01313
fragment_chain:1"

misc_feature

97779..101129
/note="assembly_fragment:00879
fragment_chain:1"

misc_feature

101230..145265
/note="assembly_fragment:00921
fragment_chain:1
clone_end:T7
vector_side:right"

BASE COUNT 44802 a 27655 c 28128 g 44480 t 200 others
ORIGIN

Query Match 3.2%; Score 73; DB 2; Length 145265;
Best Local Similarity 43.3%; Pred. No. 0.003;

Matches 340; Conservative 0; Mismatches 445; Indels 0; Gaps 0;

QY 1323 GAAACAAGTACCTCCAGACTTTACACAGAGAGTAATAATACACTCATATTGCAAGTCG 1382
DB 32864 GATATTATATTTCTTTAAAGTATACCTAATTAATATATATATATATATATATATAT 32923
QY 1383 TGACCTCTTAAATAGTGTGTAACCAAGATACGATCCGACACTTCTTAAACA 1442
DB 32924 TAAT 32983
QY 1443 TATCAAAAAAGTAAATGAGAAGGTTACAGGAAAAAGCAAGCTATGAGTATAGTG 1502
DB 32984 TAATGAT 33043
QY 1503 TCTAAGTACACACATTTGGCGGCTACTCAGTTAGCAATATATTATTCTCAGTAG 1562
DB 33044 TAAT 33103
QY 1563 TGCTGATTTAGATAGATTAATAAGACTATCATGTTTGGAGACATGAATGATAG 1622
DB 33104 TAAT 33163
QY 1623 TACTTAGAGTTGCTAAATCCTGTAGATACGCTCAAGATAGTATATCTCCACAGCT 1682
DB 33164 TAATGATGATGAT 33223
QY 1683 AACTGACCTTGATTTCTTATTTCCGATTAACAATTAATATCAATCTTATTGGAAC 1742
DB 33224 TAAT 33283
QY 1743 GTGCGATCCAGAGATTTGTTGATTTATTCGTATGAGAGATTAATAAGAACTTATAC 1802
DB 33284 TAAT 33343
QY 1803 TGTACTCATTAATTTACATTGAGAAAAAGGTGACTGTTTACGCTGAGCAACTAA 1862
DB 33344 TAATATGATGATGATATATATATATATATATATATATATATATATATATAT 33403
QY 1863 AGATTTCATTTGAAATTTGAATTAATAAGCAAGAAATTCCTTTCGAAACTGT 1922

DB 33404 TAAT 33463
QY 1923 TAAACAGATTAACCAACCTCGAATTTAAGTGTAAAGCAACCATTAATTAACA 1982
DB 33464 TAATGATGATGAT 33523
QY 1983 TGGGAAAGTTTAACTTCAAGTTTACAGAGGTTTATTTCTTACCTTGTCAAGAAAC 2042
DB 33524 TAAT 33583
QY 2043 AGATTTCGAGGCTATAGGTTAAAGTTAATAGCCAAAGATGCAATGCTACAGTTTC 2102
DB 33584 TAATATGATGATGATATATATATATATATATATATATATATATATATATAT 33643
QY 2103 AAAAA 2107
DB 33644 TAATA 33648

Search completed: August 19, 2003, 12:03:45
Job time : 8158 secs